Vol. XXIV APRIL, 1940 No. 2

THE JOURNAL OF APPLIED PSYCHOLOGY

Edited by
JAMES P. PORTER, Ohio University
Athens, Ohio

TABLE OF CONTENTS

PAUL S. ACHILLES. Report on the Activities of the Psychological Corporation—1939 109

ANDREW W. BROWN. The Development and Standardization of the Chicago Non-Verbal Examination, Part II 122

MARGARET SEDER. Vocational Interests of Professional Women, Part I 130

RALPH G. ECKERT AND NOEL KEYS. Public Speaking as a Cue to Personality Adjustment 144

PERCY W. COBB. The Limit of Usefulness of Accident Rate as a Measure of Accident Proneness 154

EDGAR A. DOLL. The Social Basis of Mental Diagnosis 160

MATTHEW LUCKIESH AND FRANK K. MOSS. Boldness as a Factor in Type-Design and Typography 170

THEODORE R. SARBIN. The Case Record in Psychological Counseling 184

ROSE G. ANDERSON. Fifth Revision of Kuhlmann-Anderson Tests 198

GEORGE SPACHE. A Comparative Study of Three Tests of Visual Acuity 207

JOHN W. HANCOCK. An Experimental Study of Four Methods of Measuring Unit Costs of Obtaining Attitude toward the Retail Store 213

HENRY C. WERNER. Comparative Study of a Small Group of Clinical Tests 231

W. L. BARNETTE. Norms of Business College Students on Standardized Tests: Intelligence, Clerical Ability, English 237

News and Notes 245

Book Reviews 247

New Books and Pamphlets 252

REPORT ON THE ACTIVITIES OF THE PSYCHOLOGICAL CORPORATION —1939—

IT is gratifying to report that through the combined efforts of the Psychological Corporation's staff and research associates over a period of years a number of services have been established and are steadily gaining public recognition. The principal activities in which the Corporation is now engaged may be briefly described as follows:

Market and Social Research—dealing with problems in the fields of selling, advertising and public opinion.

Industrial and Personnel Research—dealing with problems pertaining to production, personnel and management.

Clinical Work—under title of the Psychological Service Center and the Child Adjustment and School Service Division —rendering psychological examining and counseling services to schools and individuals on educational, vocational and personal problems.

Testing Service for Schools of Nursing—affording facilities to schools of nursing throughout the country for testing applicants for admission and for participating in selection research.

Test Division—serving as a publishing, sales, advisory, and research agency for psychological tests and materials.

Primarily concerned, of course, in the conduct of the above services are the psychologists comprising the Corporation's staff in its New York and Chicago offices. The combined staff now numbers fourteen psychologists holding the Ph.D. degree

and six, the M.A., representing among them specialized training in practically all fields of applied psychology. Many of the members have also had training in other fields, such as economics, nursing, education, or engineering, as well as practical business experience. Thus the staff compares favorably both in size and training with the faculties in psychology at many universities, and the entire force of the two offices, including statistical, secretarial and clerical assistants, numbers some forty-five persons.

Of equal importance with the staff for the successful operation of several of the Corporation's services are the many psychologists elsewhere throughout the country who cooperate with it as research associates. In 1939 over two hundred psychologists and probably over a thousand of their assistants and students participated in various ways in the activities of the Corporation. Market research work and public opinion studies involving personal interviewing constituted the chief source of assignments. However, some ninety-two psychologists conducted examinations in various localities for the Testing Service Division for Schools of Nursing, and inquiries from individuals concerning personal consultations were referred to a considerable number of psychologists prepared to undertake private consulting work.

The psychologists who are listed as research associates of the Corporation are not its employees or business agents, but serve in the capacity of independent contractors to whom the Corporation gives preference in offering assignments of work to be executed for it, or in making referrals. Obviously it is not feasible for the Corporation to extend its responsibilities beyond specific assignments. Hence its name may not be used by research associates as signifying sponsorship of their own activities or as a means of obtaining business for themselves. Nor are they authorized to commit the Corporation in any way. However, in any instance where it may be advantageous for one or more research associates to work for a client under the sponsorship of the Corporation and in cooperation with the

members of the Corporation's staff, it is entirely feasible to make suitable arrangements for this in advance through the Corporation's offices. Instances thus far in which such arrangements have been made for work in the name of the Corporation have proved mutually satisfactory, and it is to be hoped that further business may be developed with and through research associates along these lines.

Although the Corporation offers remuneration wherever possible for the execution of assignments on a contractual basis, many research associates have from time to time taken part in experimental studies for developmental or promotional purposes where no immediate return could be offered. This form of voluntary contribution of services was envisaged by the Corporation's founders as essential to the achievement of its objects and is hereby gratefully acknowledged. It accounts in large measure for the Corporation's present degree of success and is a tribute to the scientific spirit of the participating psychologists. Each year the Corporation has spent a growing portion of its gross income for research projects of its own initiation and for its internship program. The purpose of such research is to improve basic techniques. To cite but one example, the Corporation last year financed a 3,000-interview study on "The Effects of the Form of the Question on Public Opinion Polls." This study has been published as a Ph.D. thesis by A. B. Blankenship, who first started work with the Corporation as an interne.

Although the necessity of the Corporation's earning its own funds limits the extent to which it may engage in such research projects, the earning power of the Corporation is gradually reaching the point where more such studies can be undertaken. Industrial concerns support their own research for new products and the improvement of processes. Why should not the same be expected from an organization endeavoring to capitalize the economic value of applied psychology? Furthermore, the research projects for which the

Corporation is directly paid are themselves often of scientific value. Thus with regard to a taste study made for a beverage company, a leading psychologist in the field of psychophysics stated that the study was one of the most important contributions ever made in the psycho-physical field. There is much of value to psychology and psychologists in meeting the requirements for applied research in the business world. Even though the problems are practical, the requirements of ingenuity, careful experimentation and dependability of findings are often more exacting than for work on theoretical problems in the psychological laboratory or the field of so-called *pure psychology*. Thus the day-to-day work of the Corporation, although involving much that is routine and nonspectacular, nevertheless plays a useful part in developing scientific methods and data as well as in extending the areas for useful work by psychologists and the earning power for the advancement of their science.

GENERAL ACTIVITIES

To an extent perhaps not generally recognized, the Corporation's offices render informational and other services both to the public and psychologists. Many mail, telephone and personal inquiries are received daily regarding tests, the services of psychologists and all manner of psychological problems. These must be courteously and promptly handled although they are frequently interruptive of more important and productive work. Much time also goes to interviewing persons seeking employment. Although the Corporation is not an employment agency, its activities lead many psychologists and others interested in allied fields to apply to it for employment or to seek advice concerning possible opportunities elsewhere.

The Corporation is frequently visited by psychologists from various parts of the country and in fact all parts of the world. Thus among the Corporation's visitors during 1939 were psychologists or persons interested in its activities from the fol-

lowing countries: Finland, Holland, Chile, Peru, Poland, Japan, South Africa, Sweden, China, Austria, Czechoslovakia, Canada and England. Correspondence was also handled from a number of foreign countries regarding tests and materials and their translation into foreign languages. Considerable time is devoted by many members of the staff to assistance in reading, editing, and advising on articles submitted for publication to the Journal of Applied Psychology.

The professional activities of the staff naturally extend beyond the immediate work of the Corporation's offices. Many articles by staff members were published during the year in psychological, technical and popular journals, and a book, "Back to Self-Reliance," by Dr. Matthew N. Chappell of the staff was published by Whittlesey House. Staff members also presented papers, participated in round tables, or took active part in the affairs of the various professional psychological associations as well as those of other organizations such as the American Management Association, the Society for the Advancement of Management, the National Office Management Association, and the Market Research Council.

MARKET RESEARCH AND SOCIAL STUDIES

It may justly be said that the work of the Corporation accounts in large measure for the increasing importance of psychologists and psychological contributions in this field. Under the leadership of Drs. Link, Kornhauser, Freiberg, McMurry and Mr. Corby, the Corporation has pioneered in the development of periodic surveys for determining social and economic trends, in fundamental research on statistical and sampling problems, in copy-testing, and in experimenting with new techniques.

During 1939 the Psychological Corporation's Brand Barometers, based on periodic nation-wide surveys and conducted over a period of seven years, were continued and improved. These surveys were made every two months and involved a total of 36,000 interviews for the year. In addition to the

data on trends supplied to subscribers, the facilities afforded by these periodic surveys were used to advantage by many other clients for obtaining information not only on products and advertising problems but on matters pertaining to public relations, radio programs, the influence of World's Fair exhibits, and other questions. In 1940 the Corporation's schedule calls for 10,000 interviews every three months sampling all economic levels and all parts of the country in proper proportion to population.

In addition to the above periodic surveys, the Corporation's New York and Chicago offices handled some fifty separate field investigations for business and other organizations totaling approximately 60,000 interviews. Among the products and problems investigated were the following: milk containers, beer and other beverages, razors and shaving accessories, stockings, sponges, cake packaging, cosmetics, flash light batteries, water softeners, anti-freeze solutions, copy-testing problems, public relations problems, magazine readership problems, effects of radio programs, and preliminary investigations on many new products.

Two investigations of particular interest and social significance deserve brief comment. One of these was for the Community Chest in a large city. It explored the effectiveness of the Chest's promotional efforts, opinions regarding the Chest and the various welfare organizations concerned, and also the attitude toward private versus governmental support of welfare agencies. An article by W. H. Wulfeck of the Corporation's staff concerning this study was published in the Journal of Applied Psychology, Vol. 23, No. 6, Dec. 1939.

The other study represented an application to personnel problems of techniques frequently used in copy-testing and consumer studies. The directors of a large organization wished to ascertain the desires and opinions of employees regarding various pension and insurance plans. The Psychological Corporation developed a method whereby the various plans were presented in outline form to office and factory em-

ployees in their homes. The persons interviewed included a preponderance of the company's employees but a sampling of employees of other companies as well. The preferences of employees for various plans were recorded without revealing the particular company involved. The results furnished the directors of the company with very valuable information which materially aided them in reaching a decision which would be satisfactory to the majority of employees.

In addition to its work for clients, the Corporation conducted two independent surveys in 1939 on a nation-wide scale concerning political, social and other questions of interest. These studies were also used for experimentation on the effects of variation in the form of the question, its position in the interview, and the training of interviewers. The findings on many questions were released to newspapers for publication, and the results were also submitted to present clients or used as promotional material to interest others in the methods and value of public opinion research.

The extent to which different research associates of the Corporation may participate in its market research work varies not only with their location, but with their own abilities and initiative in training competent interviewers and their interest in cooperating with the Corporation in such work. In 1939 the remuneration to research associates and their students varied from a hundred dollars or less in some localities to several thousand dollars in others where the Corporation not only had frequent opportunities to make assignments but the research associates were also alert in initiating local projects to be conducted in cooperation with the Corporation. This field, although a highly competitive one, presents splendid opportunity for the application of psychological knowledge and methods. It is to be hoped that an increasing number of research associates will develop local services for business concerns, broadcasting stations, newspapers and other agencies. If this is done, local data from various areas may eventually be pooled through the Corporation's offices as well as compared for geographical or other differences. If psy-

chologists desire, however, to gain a position of leadership in this field it cannot be too strongly emphasized that the work of the Corporation in cooperation with research associates and their students must be thoroughly dependable and superior in quality. The situation is one in which the conscientious and faithful work of many psychologists working jointly can produce results of far-reaching significance. The work of individuals may seemingly be lost in the whole, but there is nevertheless ample opportunity for all research associates who are interested to make experiments and individual contributions.

INDUSTRIAL AND PERSONNEL RESEARCH

This field at present offers less opportunity for cooperative work among psychologists than does market research, but it may eventually lend itself to joint operations between the Corporation's staff and qualified research associates elsewhere for certain types of work. In 1939 clients of the New York and Chicago offices of the Corporation included two insurance companies, a coke and gas plant, a chemical manufacturing concern, an automobile company, several manufacturing plants, and several retail concerns. For two organizations, members of the Corporation's staff conducted psychological examinations and interviews with a considerable number of their recently employed young men and junior executives. This work combines the counseling of individuals as to their abilities, traits and interests, with reports to the management of their respective concerns for aid in placing and developing the men.

In a manufacturing plant for aeroplane accessories and other mechanical equipment, an unusual opportunity was afforded the Corporation for developing tests and other procedures for the selection, placement and training of employees. A "testing and research laboratory" was established in a portion of the top floor of the plant and arrangements made whereby from two to four hours testing and interviewing time was allowed for practically every employee, including many of the executives. This program has resulted in the

development of an effective battery of tests for selecting applicants with mechanical and other special aptitudes as well as in the better placement and utilization of the abilities of many of the company's present employees.

For a chemical manufacturing concern a general qualifying examination for all factory employees was prepared. This examination, made up in four alternate forms, embodied many features of psychological tests but utilized terms and content material especially pertinent to the types of jobs in the chemical manufacturing field. Similarly work was begun in the latter part of the year on the construction of tests especially adapted to the needs of an insurance company for the selection of its clerical employees.

A significant feature of the work in a coke and gas plant was the success of the Corporation's staff members, as outside consultants, in obtaining the interest and full cooperation of its unionized employees in the studies made. These studies involved job analyses throughout the plant, the preparation of an operating and training manual, and the setting up of selection procedures. Employees participated in much of the work and every step was cleared through the representatives of the union as well as those of the management. With regard to the work the Plant Superintendent subsequently wrote in an article entitled "Job Analysis: A Missing Link"—

> "Our experience indicated the wisdom of having job analyses made by competent outsiders working with the regular operators. Without the time and help given by the twenty-four regular operators in the six departments this work could not have been done in its present form.
>
> The benefits derived from the foregoing may be summarized in order of relative importance somewhat as follows: 1. inspiring confidence, 2. stimulating interest, 3. providing job and safety information, 4. supplying examination material, 5. providing a test of the accuracy of the employee manual, 6. producing a better understanding of the coordinated operation of the whole plant, 7. enhancing the appreciation of the importance of the individual in his job, 8. describing the

qualifications of the man necessary for the job, 9. exchanging ideas relative to the job, 10. broadening the education of the supervisors, 11. recording departmental vocabulary and vernacular, and 12. furnishing a sound base for a program for the selection and training of plant operators."

TEST DIVISION

Sales of the Corporation's Test Division amounted to over $65,000 for the year representing an increase of about 25% over 1938. Analysis of some 1000 sales, as a sample, shows them to be distributed approximately as follows:

Federal and State Governments	8.5%
City School Systems	18.9%
Colleges, Universities and Business Schools	39.1%
Hospitals, YMCA's, and Welfare Organizations	10.7%
Business Organizations	13.0%
Individuals	9.7%

During the year four new group tests and two books were published: The Occupational Orientation Inquiry by G. A. Wallar and S. L. Pressey, the Columbia Vocabulary Test by Irene Gansl and H. E. Garrett, the Diagnostic Reading Record by Gertrude Hildreth, and a Stenographic Aptitude Test by G. K. Bennett. The books were a revision of Gertrude Hildreth's "Bibliography of Mental Tests and Rating Scales" which brought this publication up to date, and a "Dictionary of Terms in Measurement and Guidance" by Earl B. South.

The Thurstone Vocational Interest Schedule was revised for machine scoring and its price reduced. An experimental form of a "Culture-Free Intelligence Test," devised by R. B. Cattell, was prepared and work started on its standardization prior to its publication and sale in 1940. The Corporation also undertook the manufacture and assembly of materials for the new "Bellevue Adult Intelligence Scale" prepared by Dr. David Wechsler. This test contains both verbal and performance material especially suited to the testing of adolescents and adults.

Although the margin of profit for the operation of this Division is small, it is becoming increasingly possible for its

Head and his assistants to engage in test research projects. Among these are the development of a new form of Mechanical Comprehension Test by Dr. G. K. Bennett and close cooperation in the research work of the Testing Service for Schools of Nursing.

For the past several years it has been the policy of this Division to limit the sale of tests and materials, insofar as possible, to persons with adequate psychological training for their proper use. Thus, when orders are received from an organization not already known to have qualified persons to use tests, inquiry is first made as to the nature of problems involved, whether a psychologist is retained and so forth. Although this policy involves some difficulties, it nevertheless seems to be gaining the general approval both of business executives and psychologists.

TESTING SERVICE DIVISION FOR SCHOOLS OF NURSING

Approximately 7350 applicants for schools of nursing were tested by this Division during 1939. This represents an increase of 28% over the previous year. These applicants were for 154 schools of nursing located in 16 different states, of which 23 schools were added to the program during 1939. Of the total number of applicants approximately 6175 were tested by members of the Corporation's staff and 1175 were tested by psychologists cooperating with the Corporation in other parts of the country.

During the year a continuous program of research was conducted in cooperation with the Corporation's Test Division. This research was directed chiefly to the development of two new tests and the revision and adaptation of others for machine scoring. The form of report made to schools for each applicant was also revised, improved in the manner of presenting test results and comments, and printed. This new form has met with the enthusiastic approval of directors of schools of nursing. With the cooperation of the New York State Board of Nurse Examiners a study is being carried on to test the validity of test predictions for success in the Licensing Examination required by the Board. Another

study is being carried on for the preparation of a suitable battery of tests for applicants to Practical Nurses Courses. Work is also being done on the preparation of an achievement test in Drugs and Solutions, one of the courses which is required in all schools of nursing.

A graduate student from Teachers College, Columbia University, made a study of some of the data in the Division's files as a contribution toward a Problems Course in Nursing Education. Considerable time was devoted by the Division's staff to the preparation of material on the psychological testing of nurses for presentation to various organizations throughout the country. The assistant director of the staff gave a two weeks course at the University of Minnesota on the Uses of Psychological Tests and Measures in Schools of Nursing.

PSYCHOLOGICAL SERVICE CENTER

As previously indicated the work of this Division includes services to schools and institutions as well as to individuals seeking counsel on educational, vocational or personal problems. Interviewing and testing procedures with individuals generally involve over ten hours time in each case as a basis for report and recommendations. In many instances further brief interviews are given from time to time as an aid to the individual in carrying out the program recommended. During the year over 200 persons utilized these services at the Corporation's offices and individual work was done with some 50 children in schools.

Members of the Corporation's staff engaging in clinical work have begun the preparation of a book of cases studied in the field of educational and vocational guidance which it is anticipated may be published in 1941.

FINANCIAL SUMMARY

The combined gross income of the Corporation's New York and Chicago Offices for 1939 was $253,518.97. Cost of materials and total operating expenses before taxes were $241,266.82. Federal, state and other taxes amounted to $6,294.71,

leaving a net of $5,957.44 which was added to the Corporation's reserves and working capital. As properly to be expected from the nature of the organization, the chief item in operating expense is remuneration to psychologists and their assistants for their work. This item, including salaries, accrued and paid commissions, and contractual work distributed among many psychologists and students as indicated above, amounted to $161,702.62. The order in which the various activities of the Corporation contributed to total income was as follows: Market Research, Test Division, Testing Service for Schools of Nursing, Industrial and Personnel Research, and Clinical Work. It is difficult to segregate expenditures for research as such, in as much as a large part of the Corporation's work is "research." However, a conservative estimate of the amount spent by the Corporation for experimental and developmental work not directly chargeable to client accounts and for its internship program would be $7,500. This represents actual cost, not remuneration to staff members, much of whose time necessarily goes to research, planning, promotional, and other work not immediately or directly productive of income.

During the greater part of 1939 the following served as staff members and principal assistants: *New York:* P. S. Achilles, Ph.D.; R. G. Anderson, Ph.D.; A. D. Arnheim, A.B.; G. K. Bennett, Ph.D.; A. B. Blankenship, M.A.; M. N. Chappell, Ph.D.; P. G. Corby, M.A.; M. B. Erb, Ph.D.; P. Gordon, M.A.; A. D. Freiberg, Ph.D.; J. R. Hochstim, Ph.D.; H. C. Link, Ph.D.; M. B. Lynaugh, M.A.; D. Manheimer, M.A.; E. M. Potts, M.A., R.N.; S. Roslow, Ph.D.; R. S. Schultz, Ph.D.; F. Sutton, M.A.; W. H. Wulfeck, Ph.D. As internes: L. C. Mead, Ph.D.; and A. W. Ayers, M.A.

Chicago: R. W. Callender, A.B.; A. W. Kornhauser, Ph.D.; R. N. McMurry, Ph.D.; M. W. Richardson, Ph.D.; S. G. Dulsky, Ph.D.

HENRY C. LINK, *Secretary and Treasurer*
PAUL S. ACHILLES, *1st Vice Pres. and General Manager*

THE DEVELOPMENT AND STANDARDIZATION OF THE CHICAGO NON-VERBAL EXAMINATION

ANDREW W. BROWN

Institute for Juvenile Research, Chicago

PART II*

B. *Age Norms*

ALTHOUGH our primary purpose in constructing the Non-Verbal Test was to devise a test for the deaf, the foreign born and other groups who have difficulty with the English language and for whom pantomime directions would be necessary, we considered that the test might also be of value for children who had difficulties with reading or in comprehending and mentally manipulating written symbols but who nevertheless would be able to follow verbal directions. This made it necessary to determine whether there would be any consistent difference between the results obtained when the test was given by pantomime directions and when it was given verbally to the same group of children. For this purpose two groups of children, equated as nearly as possible, were selected. To one group the test was given by pantomime directions, and to the other the test was given by the verbal directions. The average raw score with its standard deviation for each test is given in Table 3. There were approximately 100 cases in each group.

While the difference between the average score for the different tests is not great (with the exception of Test 1), it is quite consistently in favor of verbal directions. Moreover,

* Part I of this article appeared in the February 1940 issue of this Journal.

TABLE 3

A Comparison of the Scores for Pantomime and Verbal Directions
(N = approximately 100)

Test	Pantomime Directions		Verbal Directions	
	Average	S. D.	Average	S. D.
1	49.7	20.6	61.8	15.7
2	5.4	2.5	5.7	1.8
3	4.9	2.5	6.2	1.6
4	5.2	1.7	5.7	1.4
5	8.6	2.1	9.0	1.1
6	4.3	1.7	4.6	1.9
7	6.5	2.6	6.9	2.5
8	12.2	4.6	13.5	4.2
9	5.8	2.4	7.1	1.7
10	41.2	11.7	40.5	9.1

the spread for each test is greater when given by pantomime. On the basis of these results it was decided that the test should be standardized for both verbal and pantomime directions.

The norms for the test for both types of directions were established on the normal, hearing child. The selection of cases upon which to standardize the test was made as follows:

The census report for 1930 was reviewed and several areas in Chicago were selected which gave a group of about the same socio-economic status as that of the state as a whole. According to this report 67.4 per cent of the population of the state are native white and 30.7 per cent are foreign-born. The areas selected for standardization had about 70 per cent native white and about 30 per cent foreign-born. The predominating nationalities represented in the foreign-born were Scandinavian, Irish, British, German, and Polish. In terms of rental asked and of home-ownership the areas were in the middle economic class.

A representative elementary school was selected in a number of these areas and the total population of each school examined. The test was also given to a number of high school groups in different parts of Chicago, to the entire graduating

class at Mooseheart,[3] and to a number of adults at the Illinois Free Employment Bureau. At the upper age levels (above fourteen years) there was no way of determining whether or not the selection was a representative one. At the lower age levels, however, it can be assumed that in so far as all the children below fourteen years are supposed to be in school, and further in so far as the schools used were in an area which, as far as could be determined from the census, was representative of the general population of Illinois, the sampling is fairly representative.

The selection of children for verbal and pantomime directions was purely random, *i.e.* in one school grade 8A was given the test verbally, grades 8B and 7A were examined by pantomime, grades 7B and 6A verbally, grades 6B and 5A by pantomime, and so on down to the second grade. In another school grade 8A was given the test by pantomime, grades 8B and 7A verbally, and so on down through the grades. Pantomime directions were not used below the second grade, consequently more groups have been given the test verbally than by pantomime.

Tables 4 and 5 give the frequency distributions of weighted scores for each age level for verbal and pantomime directions. Table 6 gives the number of cases, the means, and the standard deviation (S. D.) for each age group and Figures 3 and 4 present these means and sigmas graphically.

Norms for the test are available in the form of mental ages, percentile ranks, and modified standard scores for each age up to 14 years. For the group 14 and above, only percentile ranks and modified standard scores will be presented.

The modified standard scores have been derived by equating them with the obtained standard scores. The mean of the modified score (M_D) was arbitrarily placed at 100, and 15 was selected as the standard deviation. The modified scores were then obtained by solving for D in the following equation:

[3] The author wishes to acknowledge the assistance of Dr. Martin L. Reymert, Director Mooseheart Laboratory for Child Research, for making this testing possible.

TABLE 4

The Frequency Distribution of Scores for Each Age Group (Verbal Directions)

Score	Age								
	6–0 to 6–11	7–0 to 7–11	8–0 to 8–11	9–0 to 9–11	10–0 to 10–11	11–0 to 11–11	12–0 to 12–11	13–0 to 13–11	14–0* and above
180–184									1
175–179							1		9
170–174							0		9
165–169							1	5	21
160–164						2	1	6	49
155–159						1	6	11	70
150–154					1	0	5	18	86
145–149					0	5	13	16	129
140–144					3	5	19	35	114
135–139					1	12	20	29	130
130–134				2	8	16	31	35	175
125–129			1	2	14	18	41	37	190
120–124			0	7	18	35	58	52	172
115–119			1	9	18	36	41	45	145
110–114			3	19	33	36	24	30	122
105–109			6	15	43	30	37	29	122
100–104			11	29	35	41	28	21	95
95–99		2	27	27	41	22	17	14	62
90–94		6	22	45	42	26	10	11	44
85–89		12	40	32	29	13	10	11	19
80–84	2	11	37	37	19	7	4	3	24
75–79	3	23	43	33	20	7	6	4	12
70–74	7	32	29	19	8	3	2	2	14
65–69	16	47	33	21	5	3	1	4	11
60–64	12	56	27	8	4	2	1	1	3
55–59	29	43	20	4	4	1	0	4	6
50–54	43	45	12	4	1	0	0		4
45–49	31	32	7	0	1	1	2		2
40–44	35	23	9	3	1	1			2
35–39	26	16	6	1	1	1			0
30–34	26	16	3	0	0				0
25–29	22	10	3	1	0				1
20–24	18	6	1		1				1
15–19	21	6	2		0				
10–14	14	6	1		1				
5–9	6	4							
0–4	7	2							
Total	318	398	344	318	352	324	379	423	1844
Mean	41.1	56.7	75.3	89.0	99.3	110.0	118.5	122.8	125.6
S. D.	17.6	18.1	18.7	17.2	18.8	18.6	18.6	20.8	22.2

* This column includes the scores for both verbal and pantomime directions.

TABLE 5

The Frequency Distribution of Scores for Each Age Group (Pantomime Directions)

Score	Age							
	7–0 to 7–11	8–0 to 8–11	9–0 to 9–11	10–0 to 10–11	11–0 to 11–11	12–0 to 12–11	13–0 to 13–11	14–0* and above
180–184								1
175–179							1	9
170–174							1	9
165–169							1	21
160–164							6	49
155–159				1	1	3	11	70
150–154				0	1	1	11	86
145–149				0	3	2	15	129
140–144				0	4	2	18	114
135–139				2	5	12	9	130
130–134				2	5	12	33	175
125–129			1	5	9	19	31	190
120–124			3	1	14	21	33	172
115–119		1	3	8	14	19	33	145
110–114		2	6	14	19	25	27	122
105–109		6	3	17	20	16	24	122
100–104		6	14	20	25	20	18	95
95–99	1	9	23	20	17	12	15	62
90–94	1	9	20	14	10	11	12	44
85–89	4	13	22	11	16	8	9	19
80–84	4	14	26	12	10	7	7	24
75–79	4	13	14	9	5	4	4	12
70–74	4	18	14	7	4	1	5	14
65–69	5	7	18	3	2	3	3	11
60–64	4	17	5	2	5	2	3	3
55–59	6	13	9	5	3	1	3	6
50–54	6	4	5	4	0	0	0	4
45–49	5	7	1	6	2	0	2	2
40–44	2	8	3	3		1		2
35–39	6	6	1	2		0		0
30–34	5	6	1	0		0		0
25–29	8	6	1	1		1		1
20–24	3	3	1					1
15–19		3	2					
10–14			1					
5–9			1					
Total	68	171	198	169	194	203	335	1844
Mean	54.6	68.9	81.8	92.3	104.3	111.1	119.1	125.6
S. D.	20.7	23.5	20.7	22.8	20.6	20.4	23.3	22.2

* This column includes the scores for both verbal and pantomime directions.

TABLE 6

The Number of Cases, Mean Score, and Standard Deviation for Each Age Level for Both Verbal and Pantomime Directions

Age	Number of Cases		Verbal		Pantomime	
	Verbal	Pantomime	Mean	S. D.	Mean	S. D.
6–0 to 6–11	318		41.1	17.6		
7–0 to 7–11	398	68	56.7	18.1	54.6	20.7
8–0 to 8–11	344	171	75.3	18.7	68.9	23.5
9–0 to 9–11	318	198	89.0	17.2	81.8	20.7
10–0 to 10–11	352	169	99.3	18.8	92.3	22.8
11–0 to 11–11	324	194	110.0	18.6	104.3	20.6
12–0 to 12–11	379	203	118.5	18.6	111.1	20.4
13–0 to 13–11	423	335	122.8	20.8	119.1	23.3
14–0 & above*	1844		125.6	22.2	125.6	22.2

* For the group 14 and above the cases for verbal and pantomime directions were combined.

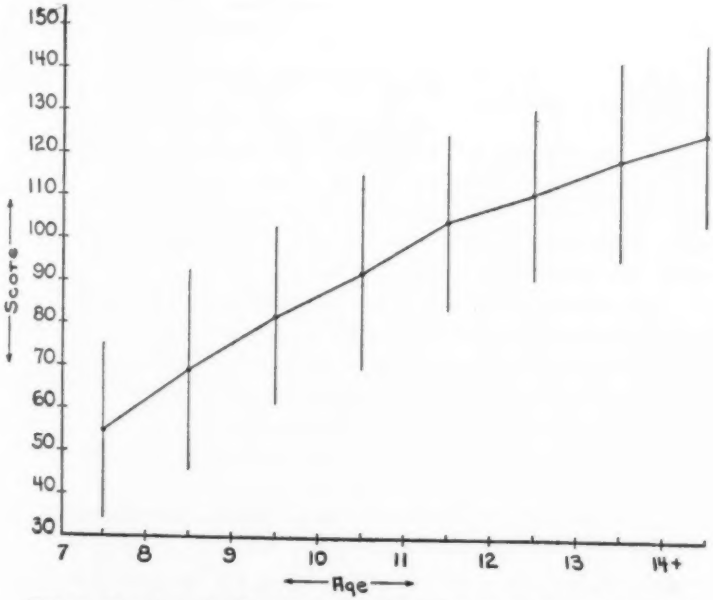


FIG. 3. Showing the Mean and Standard Deviation at each age level (Verbal Directions).

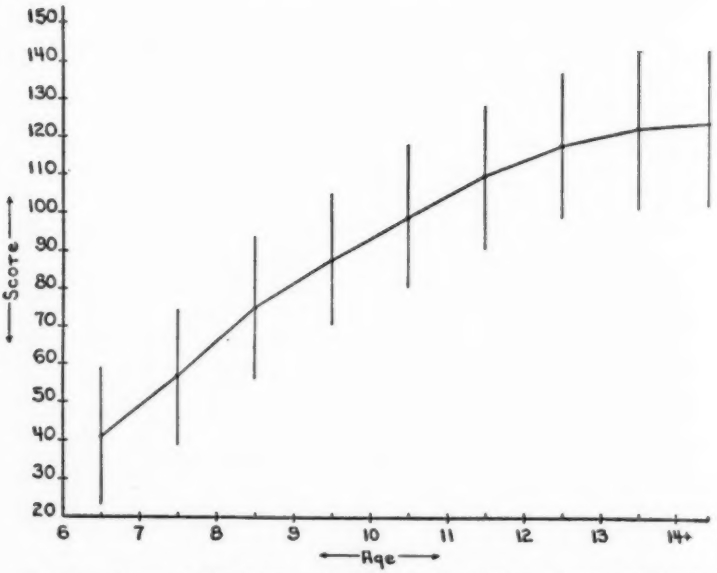


FIG. 4. Showing the Mean and Standard Deviation at each age level (Pantomime Directions).

$$\frac{D-M_D}{\sigma_D} = \frac{X-M_X}{\sigma_X}$$

where D is the modified standard score, X the raw score and M_D, M_X, σ_D and σ_X the means and sigmas of the modified scores and the raw scores respectively.

This method of presenting standard scores has the advantage of making them comparable to intelligence quotients and thus making them more acceptable in clinics, schools and courts than if they were presented in the unmodified form. If we assume a normal distribution this method would give the following proportions for the various scores and sigma limits. (See p. 129.)

These are approximately the same proportions as are given by Merrill[4] for the same IQ limits.

[4] Merrill, Maud A., "The Significance of IQ's on the Revised Stanford-Binet Scales." The *Jr. of Ed. Psychol.*, Dec. 1938, pp. 641–651.

S. S. Limits	Sigma Limits	Per cent	Descriptive Classification
130 and over	+2.00 and above	2.33	Very Superior
120–129	+1.33 to +1.99	7.01	Superior
110–119	+ .67 to +1.32	16.44	High Average
90–109	± .66	48.44	Average
80–89	− .67 to −1.32	16.44	Low Average
70–79	−1.33 to −1.99	7.01	Borderline
0–69	−2.00 and below	2.33	Mentally Defective

Modified standard scores have been computed for the raw scores at each age level and have been placed on a chart for ease of calculation.

SUMMARY

1. Age norms have been established on a test which purports to measure general intelligence without the use of verbal symbols.

2. The test has been standardized for both verbal and pantomime directions.

3. With verbal directions the test can be applied to children from age 6 through the adult level; with non-verbal directions, to children from age 7 through the adult level.

4. The test-retest and odd-even reliability for both verbal and pantomime directions is between .80 and .90.

5. The test satisfies fairly well a number of the usual criteria used in establishing the validity of intelligence tests.

6. The test has been used clinically at the Institute for Juvenile Research for more than a year and has been found of value for children with a language difficulty. The impression resulting from the use of the test in the clinic is that the ratings derived from it are, in comparison with Stanford-Binet ratings, somewhat too low at the lower levels of the IQ range and somewhat too high at the upper levels of the IQ range.

VOCATIONAL INTERESTS OF PROFESSIONAL WOMEN*

Part I

MARGARET SEDER
The Educational Records Bureau

INTRODUCTION

THE vocational interests of women have received less attention than those of men but from the beginning of the history of measurement of interests, it was assumed that women's interests were different from those of men. For example, Strong says about his Vocational Interest Blank, "It is very questionable whether women should be scored on this test. They should not be given their ratings unless an emphatic statement is made that no one knows the significance of their ratings" (3).

There have been, up to the present time, only two attempts to study the interests of women in comparison with the interests of men. When Terman and Miles devised their masculinity-femininity test, they included items sampling a wide variety of interests, and more than a third of these items were taken from the Strong Vocational Interest Blank (5). They found that the test of interests and attitudes was among the most valuable in the battery. Women were found to prefer indoor, artistic and decorative activities while men were found to prefer activities involving adventure, bodily risk, and muscular strength. The same trend of differences in interest was found by Carter and Strong, who devised a mascu-

* These data were taken from "Vocational Interest Patterns of Professional Women," by Margaret Seder, a Ph.D. thesis on file in the University of Minnesota Library. The study was carried out under the direction of Professor Donald G. Paterson.

linity-femininity scale for the Strong Vocational Interest Blank (1). The interests of the sexes are enough different that it was possible to devise a scoring key on which only three per cent of men are more feminine than the average woman and no woman is more masculine than the average man.

The fact that vocational interests are distributed differently among men and women in the general population seems well-established. Strong found significant differences between mean scores on the various occupational keys for such groups as husbands and wives, high school boys and high school girls who were matched for age, grade and intelligence, and college men and women matched on the same bases (2). However interesting these differences may be, they shed no direct light on the question, important for vocational guidance, of whether men and women engaged in the same occupation have the same interests or whether there are characteristic sex differences in interests within a given occupation.

Without formulating or attempting to answer this question, Strong has composed a Vocational Interest Blank for Women with scoring keys for seventeen occupations and masculinity-femininity. Whether whatever these keys are measuring is anything different from the occupational interest patterns measured by the keys for the men's blank was not determined. Strong says, "Occupationally speaking, they (women) will seek somewhat different careers; and within any occupation or profession, it is most likely that they will emphasize different aspects of those occupations. For not only do women have, to a large degree, different interests from those of men, but those same differences constitute one of the general factors in terms of which the occupational interests of men alone are to be explained" (4). No evidence is cited in support of this claim that the interests of women will differ from those of men in the same occupation.

The present study was undertaken, therefore, to determine whether the keys for the Strong Vocational Interest Blank for Women were giving measures of anything different from the men's blank, that is, whether the interests of women dif-

fered from those of men engaged in an occupation of the same name.

METHOD

The two vocational interest blanks used in this investigation were the Vocational Interest Blank and the Vocational Interest Blank for Women, both by E. K. Strong, Jr. Analysis of items on the blanks showed that 259 items were common to both blanks and 19 additional items were similar. The men's blank contained 142 items which did not occur on the women's blank, and the women's blank contained 132 items not appearing on the men's blank. The men's blank was scored for 25 occupations and masculinity-femininity, and the women's blank was scored for 16 occupations and masculinity-femininity. The nine occupational keys common to both blanks are dentist, physician, artist, lawyer, life-insurance sales, teacher, Y.M.C.A.-Y.W.C.A. secretary, office clerk, and purchasing agent.

The two blanks were filled out by each of 200 women, 100

TABLE 1

*Statistics Descriptive of Age, Education and Experience of the Subjects of This Investigation and of the Original Criterion Groups**

	Age		Experience		Education	
	Mean	σ	Mean	σ	Mean	σ
100 Life Insurance Saleswomen	46.2	9.1	7.1	5.1	12.6	2.5
Criterion Group of Life Insurance Saleswomen (N = 77)	46.9		3 or more years		13.5	
100 Women Physicians	44.1	12.0	15.9	10.5	M.D. degree or more	
Criterion Group of Women Physicians (N = 196)	41.0		3 or more years		17.4	

* Taken from the manual for the Vocational Interest Blank for women (4).

life-insurance saleswomen and 100 women physicians. These groups were selected as being fairly representative of business women and of professional women, respectively. All subjects were secured by sending letters requesting cooperation to women in middle western cities. About one-third of the women who were asked to fill out the blanks responded by returning both blanks completed.

Table 1 shows the statistics descriptive of the age, education, and experience of the groups and similar data for the criterion groups of women on whom the keys for these two occupations were based. The groups of this study were probably less highly selected than the criterion groups. There was, for example, no minimum set as to the amount of insurance sold or the length of experience of the saleswomen. These physicians are slightly older than the criterion group. Strong stipulated that none included in his criterion groups should be over 59 years of age and none should have less than three years' experience. Eight of the saleswomen and 14 of the physicians were older than 59 years, and 18 saleswomen and 3 physicians had had less than 3 years' experience. All subjects of this study were actively engaged in their occupations at the time they filled out the blanks.

Additional data bearing on the working hypothesis were obtained by analyzing the nine pairs of occupational keys for the two blanks. Since the two blanks contain 268 identical or comparable items and since the weightings for each item are based on actual responses made by professional men or women, comparison of the interests of men and women in the criterion groups was accomplished by comparing item weights. These data are reported in Part II, to be published in a later issue of this journal.

RESULTS

Three types of analysis of the data obtained from the physicians and saleswomen were made to determine whether the data tend to substantiate or to refute the working hypothesis, which is that the interests of men and women engaged in a

given occupation are the same. The first was an examination of the number of A, B, and C ratings obtained by each of the two groups in the nine occupations scored for both blanks; the second was a consideration of the correlations between scores for these occupations scored for both blanks in comparison with reliability coefficients cited by Strong; and the third was a factor analysis of the intercorrelations of scoring keys for each occupational group of subjects.

The first step in analyzing the data was to see whether the women physicians got higher ratings when compared with women physicians than when compared with men physicians and whether life insurance saleswomen got higher ratings when compared with saleswomen than with salesmen. The data are presented in Tables 2 and 3. These ratings are of special interest because, for counseling purposes, ratings rather than

TABLE 2

Number of A, B, and C Ratings Obtained by 100 Physicians in Nine Occupations Scored for the Two Vocational Interest Blanks

	Men's Blank			Women's Blank		
	A	B	C	A	B	C
Dentist	10	65	25	25	52	23
Physician	62	33	5	58	36	6
Artist	5	37	58	18	45	37
Lawyer	23	71	6	15	45	40
Life Insurance Sales	0	20	80	0	9	91
Teacher	30	46	24	19	32	49
Y.M.C.A.-Y.W.C.A. Secretary	4	28	68	2	31	67
Office Clerk-Stenographer-Secretary	2	30	68	1	27	72
Purchasing Agent-Office Worker	0	15	85	3	40	57

numerical scores are used. An A rating is said to show close agreement with the interests of the criterion group, and for the men's keys it includes 75 per cent of each criterion group, the dividing line being placed at $-1\,Q$; and for the women's keys, it includes 69 per cent of each criterion group, the dividing line being placed at $-.5\,\sigma$. B (some agreement) is the

TABLE 3

Number of A, B, and C Ratings Obtained by 100 Life Insurance Saleswomen in Nine Occupations Scored for the Two Vocational Interest Blanks

	Men's Blank			Women's Blank		
	A	B	C	A	B	C
Dentist	0	21	79	0	30	70
Physician	2	26	72	1	27	72
Artist	1	8	91	6	21	73
Lawyer	10	76	14	18	51	31
Life Insurance Sales	44	41	15	34	33	33
Teacher	12	48	40	7	31	62
Y.M.C.A.-Y.W.C.A. Secretary	10	49	41	6	41	53
Office Clerk-Stenographer-Secretary	15	53	32	16	55	29
Purchasing Agent-Office Worker	5	39	56	24	61	15

rating assigned to the 24 per cent of men in each criterion group scoring between – 1 Q and – 3.5 Q, and to the 30 per cent of women in each criterion group scoring between $-.5\,\sigma$ and $-2.5\,\sigma$. The remaining one per cent of the distribution of each criterion group is assigned the rating of C (very little agreement). Because the ratings do not mean exactly the same thing for both blanks, small differences in the number of women scoring A's on any pair of keys must not be over-emphasized. The similarity in number of ratings assigned by the two keys should be noted especially for the Physician, Life Insurance Sales, Y.M.C.A.-Y.W.C.A. Secretary, Office Clerk and Purchasing Agent keys for the 100 physicians, and for the Dentist, Physician, Artist, Teacher, Y.M.C.A.-Y.W.C.A. Secretary, and Office Clerk for the 100 saleswomen. Physicians appear to resemble women dentists and artists more than men in these professions, and men lawyers and teachers more than women. The saleswomen appear to resemble women lawyers and purchasing agents more than men in these professions and life insurance salesmen more than women.

The danger of comparing ratings on the men's and women's blanks is illustrated here. If we count the number of sales-

women scoring as high as the top 75 per cent of criterion saleswomen, there are 43, as compared with 44 who score as high as the top 75 per cent of salesmen. Of the 100 physicians, 62 score as high as the top 75 per cent of the criterion group of men physicians, and 63 score as high as the top 75 per cent of women physicians in the criterion group.

These data offer no evidence for important differences between interests of men and women physicians nor between the interests of life insurance salesmen or women. Of course, if other groups were to be tested, marked differences might be found. For these groups of women, however, the men's blank would serve about as well as the women's blank.

Scores on these nine pairs of keys were then correlated to see how much agreement there was between numerical scores derived from the men's blank and those derived from the women's blank. Scores on the two masculinity-femininity keys were also correlated. The correlations, together with the odd-

TABLE 4

Correlations Between Occupations of the Same Name Which are Scored for Both Blanks, and the Reliabilities of Those Keys for the Women's Blank

Key	r based on 100 physicians	r based on 100 life insurance saleswomen	Odd-even r* of women's key
Dentist	+.50	+.60	+.78
Physician	+.68	+.55	+.87
Artist	+.71	+.71	+.93
Lawyer	−.03	+.06	+.81
Life Insurance Sales	+.55	+.72	+.74
Teacher	+.71	+.57	+.79
Y.M.C.A.-Y.W.C.A. Secretary	+.80	+.76	+.88
Office Clerk-Stenographer-Secretary	+.65	+.76	+.85
Purchasing Agent-Office Worker	+.54	+.55	+.92
Masculinity-Femininity	+.77	+.57	+.74

* Taken from Strong (4).

even reliability coefficients for the women's keys, are listed in Table 4.

The most striking thing about the data in this table is the consistently good-sized correlations except for those for the lawyers' keys where there is no relationship at all. It is possible that the two groups of subjects used had interests of such a nature that the difference with regard to the lawyer keys was emphasized. It is probably more reasonable to assume that there were real differences between interests of the criterion groups of men and women lawyers.

The correlation between physician keys for physicians is +.68 and for life insurance saleswomen is +.55. The correlation between life insurance keys for the saleswomen is +.72 and for the physicians is +.55. Both times the higher correlation is based on the group in that occupation. This suggests that the other correlations might be higher if they were based on a group of the same occupation. Nevertheless, none of them are very much lower than the odd-even reliabilities reported for the women's blank. Except for the lawyer keys, the lowest correlations are for the purchasing agent-office worker keys, and other evidence indicates that this is due at least partly to selection of the criterion groups.

The rest of the correlations between keys are about of the magnitude of alternate form reliability coefficients when compared with the odd-even reliability coefficients cited. In several instances, there is probably not a significant difference between the obtained correlation and the odd-even coefficient. This is true, for example, of life insurance sales keys based on data from the saleswomen and of the teacher keys based on data from the physicians.

Since these first analyses gave little indication of sex differences between men and women engaged in the same occupations, factor analyses were undertaken to determine whether sex differences might be manifested in either or both of two ways. It was thought that the factor loadings for some of the pairs of occupations might be different or that some factor might show heavy loadings for keys for one blank and not for

keys for the other blank. Thirty-five keys were used for the factor analyses, and in all of this work also, the two groups of subjects were kept separate. The keys used included all keys for the women's blank and 18 of the 26 scored for the men's blank. In addition to the ten keys common to both blanks, the eight men's keys added were engineer, personnel manager, chemist, journalist, psychologist, real estate salesman, mathematician, and physicist.

The factor loadings for the ten keys common to the two blanks are listed in Tables 5 and 6 together with the com-

TABLE 5

*Centroid Factor Loadings for 20 Vocational Interest Keys, Based on Scores of 100 Women Physicians**

Key	Factor 1	2	3	4	5	Communality
1	−.38	+.41	−.19	−.35	+.09	.47
2	−.83	+.16	+.05	−.26	+.15	.81
3	−.62	−.27	+.37	−.51	−.15	.87
4	−.55	−.56	+.07	+.11	+.33	.74
5	+.13	−.64	−.41	−.16	+.52	.88
6	+.66	+.41	+.47	−.10	+.22	.88
7	+.90	+.14	+.24	−.20	+.32	1.03
8	+.77	+.37	−.32	−.04	−.05	.83
9	+.07	+.23	−.73	+.25	−.26	.72
10	−.14	+.49	−.29	+.75	−.17	.86
1f	−.28	+.57	−.50	−.07	+.20	.71
2f	−.76	+.10	+.08	+.34	+.30	.80
3f	−.84	−.25	+.03	−.09	−.15	.79
4f	+.36	−.15	+.14	+.78	+.14	.79
5f	+.25	−.37	−.40	+.32	+.28	.54
6f	+.51	+.30	+.65	+.07	−.14	.79
7f	+.77	+.08	+.32	−.03	+.26	.76
8f	+.71	−.02	−.40	−.14	−.40	.85
9f	+.67	+.30	−.48	−.16	−.30	.88
10f	−.40	+.35	−.25	+.58	+.18	.72

* The factor loadings and communalities were computed to 4 decimal places. They are reported as 2 place numbers.

munalities (sum of squares of factor loadings for each test).* The first ten keys are for the men's blank and are in order (1)

TABLE 6

Centroid Factor Loadings for 20 Vocational Interest Keys, Based on Scores of 100 Life Insurance Saleswomen

Key	Factor 1	2	3	4	5	Communality
1	+.10	−.44	−.56	+.38	+.36	.79
2	−.41	−.60	−.43	+.33	+.22	.87
3	−.69	−.42	−.11	+.38	+.22	.86
4	−.82	+.08	+.14	+.09	−.28	.78
5	−.24	+.80	+.24	+.07	+.21	.81
6	+.60	−.41	+.50	+.32	−.07	.88
7	+.64	−.13	+.66	−.14	+.43	1.06
8	+.79	+.43	−.24	+.25	−.32	1.04
9	+.45	+.39	−.33	−.14	−.26	.54
10	+.34	−.20	−.30	−.75	−.15	.83
1f	+.43	−.25	−.54	+.10	+.28	.63
2f	−.32	−.61	−.21	−.34	+.08	.63
3f	−.81	−.26	−.30	−.22	+.17	.90
4f	+.11	+.07	+.52	−.49	−.27	.60
5f	−.24	+.66	+.15	−.33	−.16	.65
6f	+.23	−.55	+.31	+.30	−.34	.65
7f	+.44	−.17	+.66	+.21	+.11	.71
8f	+.64	+.58	−.05	+.24	+.03	.81
9f	+.75	+.51	−.14	+.28	−.14	.95
10f	+.22	−.07	−.29	−.42	−.26	.39

dentist, (2) physician, (3) artist, (4) lawyer, (5) life insurance salesman, (6) teacher, (7) Y.M.C.A. secretary, (8) office clerk, (9) purchasing agent, and (10) masculinity-femininity. Keys numbered 1f to 10f are for the women's blank and are arranged in a corresponding order.

The first thing to be noticed is how the loadings for similar keys for the two blanks, 1 and 1f, 2 and 2f, and so on, compare for any one factor. The difficulty encountered immediately

* The complete tables of factor loadings and the tables of intercorrelations are to be found in the thesis, op. cit.

is that the sampling distribution of loadings is not known, so it is not possible to test for the significance of differences between loadings nor between communalities. It is necessary simply to use an inspection method. Some idea of the magnitude of differences to be expected may be had by comparing the factor loadings for the two M–F keys (10 and 10f), since we know that these keys are very similar, the correlation between the M–F key for the women's blank, which is based only on items common to the two blanks, and this key for the men's blank, which is based on all items, being +.87 (4). The largest difference between the M–F key for the men's blank and the same key for the women's blank is .35 based on the physicians, for the fifth factor, and .33 based on the insurance saleswomen, for the fourth factor. It is, of course, quite likely that units are not equal, and if this is so, perhaps we may use the squared factor loading—the factor's contribution to the test's communality—instead. With the physicians, the loadings for the fifth factor being +.18 and −.17, the difference, $(X^2 - Y^2)$ is .0035. With the insurance saleswomen, whose loadings for the fourth factor are −.75 and −.42, the difference is .3861. Perhaps, therefore, we may expect differences as great as .4 between the contributions of a factor to two tests' communalities by chance. Such speculation may be quite futile. It should be pointed out here that both of these large differences come from the last two factors where it seems probable that chance errors in the original correlations are having a greater influence than in earlier factors.

The pairs of values of most interest for each group are its own occupational keys. For physicians the pairs of factor loadings (Table 5, keys 2 and 2f) are −.83 and −.76, +.16 and +.10, +.05 and +.08, −.26 and +.34, +.15 and +.30. The largest discrepancy is the one for factor four and it is possible that neither loading differs significantly from zero. The communalities are practically identical, +.81 and +.80. For insurance saleswomen the loadings (Table 6, keys 5 and 5f) are −.24 and −.24, +.80 and +.66, +.24 and +.15, +.07 and −.33, and +.21 and −.16. The communalities are not so simi-

lar this time, + .81 and + .65, but taken as a whole, these values might lead one to conclude that the keys based on men and women in the same occupation were essentially the same.

It is interesting to notice that the agreement of factor loadings for physicians' keys based on the data from the saleswomen (Table 6, keys 2 and 2f) is of substantially the same order as that found with the physicians although the communalities are not so similar. Again the largest discrepancy is for factor four, where the loading for the men's key is + .33 and the loading for the women's key is – .34. The agreement of factor loadings for the insurance keys based on the data from the physicians (Table 5, keys 5 and 5f) shows the largest discrepancy to be for factor four, as in the analysis based on the saleswomen. In this case as well, the difference between communalities obtained from data from the other group (physicians) is greater than that obtained with the life insurance group on the insurance keys.

The largest discrepancies for any of the nine keys are found for the lawyers' keys, 4 and 14, for both factor analyses, and with these keys in each case the largest differences are for factors 1 and 4. When we go back to the original correlations (see Table 4) we find them to be – .03 (physicians) and + .06 (saleswomen) so we should expect to find very little agreement among factor loadings.

The second point on which the factor analyses were to throw light was whether any factor showed heavy loadings for the men's keys only or for the women's keys only. Inspection of Tables 5 and 6 shows that the range of loadings is about the same for the men's and women's keys for any one factor. That is, it would not be possible on the basis of the loadings for any factor to separate the keys into two groups, those based on women and those based on men. Loadings for the other fifteen keys included in the factor analyses are not listed here, but they follow exactly the same trend as the loadings included in the tables.

It is true that the factors listed are centroid, not "psycho-

logical'' factors, and for purposes of naming the factors, it might be necessary to rotate axes. It is quite beside the point to try to determine here what these factors are. The fact that they exist in both men's and women's interests is the important one. It might be possible by rotation of axes to get approximately the same factor loadings for the two groups of subjects, but again, since there is no test for the significance of differences between loadings the task seems useless. However, communalities are invariant of rotation, so one interested in the degree of similarity obtained from the two correlation matrices, may examine the communalities. Lacking tests of significance, we simply point out the similarity for variables 2, 3, 4, 5, 6, 7, 10, 1f, 7f, 8f, and 9f where the difference between communalities obtained for the two analyses is less than .1, and the relatively greater difference for variables 1, 8, and 10f where the difference is greater than .2.

If it is not possible to compare factor loadings for the keys from one group to the other, it is possible to get some idea of the correspondence of data between these women physicians and saleswomen by transforming the original intercorrelations into Zr's and then correlating the corresponding Z transformations. The resulting correlation, based on the 595 pairs of intercorrelations, is + .89. Most of the differences, if standard errors of the differences between pairs were computed, would be accounted for on the basis of chance variation.

SUMMARY

1. Both the women physicians and the life insurance saleswomen obtained much the same ratings on the men's blank as on the women's blank for occupations for which both blanks were scored. The agreement is particularly close when the physicians are compared with the top 75 per cent of the criterion men and women physicians and when the saleswomen are compared with the top 75 per cent of criterion salesmen and women.

2. The correlations of interest scores on the men's keys with

scores on women's keys of the same occupational name are of the magnitude of alternate form reliability coefficients as compared with odd-even reliability coefficients, except in the case of scores on the lawyers' keys and perhaps in the case of scores on purchasing agent-office worker keys.

3. The results of factor analyses show that keys with the same names for the two Vocational Interest Blanks have similar factor loadings except in the case of the lawyers' keys.

4. No factor appears with heavy loadings just for the men's keys or just for the women's keys.

BIBLIOGRAPHY

1. Carter, H. D. and Strong, E. K. Jr. "Sex Differences in Occupational Interests of High School Students," *Personnel Journal*, XII, 1934, 166–175.
2. Strong, E. K. Jr. "Interests of Men and Women," *Journal of Social Psychology*, VII, 1936, 49–67.
3. ————. *Manual for the Vocational Interest Blank*, Stanford University Press, 1935.
4. ————. *Manual for the Vocational Interest Blank for Women*, Stanford University Press, 1936.
5. Terman, L. M. and Miles, C. C. *Sex and Personality*, New York, McGraw-Hill, 1936.

PUBLIC SPEAKING AS A CUE TO PERSONALITY ADJUSTMENT[1]

RALPH G. ECKERT
San Jose State College

AND

NOEL KEYS
University of California

IS effectiveness as a speaker significantly related to social and emotional adjustment, and to what extent is it indicative of such "traits" as sociability, self-confidence, or neuroticism? Two studies published in 1934 have dealt with the ability of hearers to appraise personality on the basis of voice alone. That by Allport and Cantril[2] had eighteen men read over loudspeaker, radio, and from behind a screen; while Taylor[3] used phonograph records from fifteen students selected as highly neurotic or the opposite. Both found a close consensus among the various auditors in the judgments expressed. In the first-named study, listeners also showed much better than chance success in associating voices heard with photographs and brief character sketches of the readers. Taylor, however, discovered that judgments as to personality agreed scarcely at all with the subjects' own self-ratings. Indeed, the consensus among auditors was highest on those judgments which were most at variance with the readers' scores on personality tests. Misleading stereotypes may therefore have accounted for much of the agreement observed.

[1] Assistance in the preparation of these materials was furnished by the personnel of the Works Progress Administration Official Project No. 11550.

[2] Allport, G. W., and Cantril H. "Judging Personality from Voice." *Journal of Social Psychology* 5: 37–55. February, 1934.

[3] Taylor, H. C. "Social Agreement on Personality Traits as Judged from Speech." *Journal of Social Psychology* 5: 244–248. May, 1934.

Stagner[4] had five men and five women read aloud before a class in social psychology. The highest correlations reported were one of .80 between ratings of the five women subjects on voice intensity and their self-confidence as measured by the Bernreuter test, and one of –.50 between ratings for poise and Bernreuter scores for sociability. These findings impressed Stagner as so inconsistent and illogical that he concluded "there is no agreement between self-judgments and social judgments based on vocal cues."

A public speaking situation would seem to afford uncommon opportunities for self-revelation. That the bearing of a typical college student before an audience of his peers should fail to indicate significant traits of personality and adjustment is difficult to believe. The impression made, however, is by no means dependent on voice alone. In the study here to be reported no attempt was made to confine the judgments to ratings based solely on auditory cues, and the subjects did not read but speak. Moreover, the numbers involved were several times as great as in previous investigations.

SUBJECTS AND PROCEDURE

Students entering San Jose State College in California are given the American Council Psychological Examination as a measure of scholastic aptitude or intelligence, and the Bernreuter Personality Inventory as an indication of social and emotional adjustment. About forty-five per cent elect public speaking as a subject of instruction in their first or second year. In the winter of 1938 advantage was taken of these facts to investigate the relation between students' characteristics as speakers and their scores on tests of intelligence and personality. For this purpose, a procedure for obtaining speech ratings was devised, and the tests taken upon admission to college were supplemented by the Bell Adjustment Inventory given to the classes concerned at the opening of term. In all, ten sections of a speech class taught by five different

[4] Stagner, Ross. "Judgments of Voice and Personality." *Journal of Educational Psychology* 27: 272–277. April, 1936.

instructors took part and one hundred fifty six individuals were fully rated. Of these, approximately one half were freshmen, with girls slightly outnumbering boys.

One month after the opening of school, each student was asked to talk for two minutes on the topic, "If I could go where I wished, I would go to . . . ." A month later each was again called upon to speak more or less extemporaneously upon a topic assigned at random from a pre-announced list of twenty simple subjects, such as "Double Feature Movies," "My Favorite Sport," and "What This College Needs Most."

Several class sections met jointly on these occasions, so as to provide an audience of reasonable size; and each speaker was rated independently by from three to four speech instructors, two of whom were the same for all. In order to minimize constant differences between raters, standard scores were used in averaging.

NATURE AND RELIABILITY OF RATINGS

Seven characteristics were selected for observation, as follows: (1) vivacity, (2) poise versus self-consciousness, (3) quality of voice (pitch and timbre), (4) expression, (5) fluency as opposed to jerky, halting delivery, (6) articulation, and (7) general effectiveness as a speaker. A graphic seven-point scale was constructed for each, and the judges instructed in its use.

Reliability of the ratings, as reckoned in Table 1, may be regarded as reasonably satisfactory, averaging .77 for the six specific traits, and .92 for the more inclusive "general effectiveness."

As shown in Table 2, intercorrelations among ratings on the seven traits average .64 as compared with a mean reliability of .79. This would suggest that, while "halo effect" has colored the ratings on individual characteristics to a considerable extent, some measure of discrimination has been preserved. Moreover, the intercorrelations, though doubtless spuriously high, seem otherwise logical. Thus vivacity correlates most highly with expressiveness (.70) and lowest with

TABLE 1

Reliability of Total Ratings as Determined by Spearman-Brown Formula (N = 156)

Trait rated	Agreement between average ratings on two occasions		Reliability of total by Spearman-Brown formula
	r	PE	
1. Vivacity	.69	.028	.82
2. Poise	.52	.039	.68
3. Voice	.67	.029	.80
4. Expression	.59	.035	.74
5. Fluency	.60	.034	.75
6. Articulation	.70	.029	.82
7. General effectiveness	.85	.014	.92
Mean for traits 1 to 6	.63		.77
Mean for traits 1 to 7	.66		.79

fluency (.49), owing to the unevenness of delivery which sometimes results. General effectiveness as a speaker seems to depend most largely upon expression (.85), and least upon pitch and timbre of voice (.54).

TABLE 2

Intercorrelation of Average Ratings on Seven Speaking Traits (N = 156)

Trait rated	1	2	3	4	5	6	7	Mean	Reliability
1. Vivacity									.82
2. Poise	.55								.68
3. Voice	.54	.53							.80
4. Expression	.70	.63	.66						.74
5. Fluency	.49	.65	.37	.62					.75
6. Articulation	.54	.65	.52	.67	.68				.82
7. General effectiveness	.76	.77	.62	.85	.77	.76			.92
Mean intercorrelation	.60	.63	.54	.69	.60	.64	.76	.64	.79

MEASURES OF PERSONALITY ADJUSTMENT

The Bernreuter and Bell Inventories are among the most widely used instruments for the appraisal of adjustment on

the college level. The former consists of one hundred twenty-five, and the latter of one hundred ten questions, each of which is answered "Yes," "No," or "?."

The Bernreuter Personality Inventory provides scales for six traits. Those chosen for the present study were as follows:

B1–N. Neuroticism. Persons scoring high on this scale are said to tend toward emotional instability. This is not only the most reliable and valid of the author's original four scales, but also substantially equivalent to his measure of introversion.

F1–C. Confidence. Persons scoring high on this tend to be too self-conscious and suffer from feelings of inferiority.

F2–S. Sociability. Persons scoring high on this supposedly tend to be unsocial, solitary or independent.

The Bell Adjustment Inventory yields scores for home, health, social, and emotional adjustments, respectively. Cor-

TABLE 3

Correlation of Intelligence Scores on the American Council Psychological Examination with Speech Ratings and Personality Inventories

	Correlation with intelligence
Ratings as speakers	
Vivacity	.21 ± .06
Poise	.34 ± .05
Voice	.16 ± .06
Expression	.30 ± .05
Fluency	.38 ± .05
Articulation	.25 ± .06
General effectiveness	.32 ± .05
Bernreuter Inventory	
Neuroticism	.00 ± .08
Lack of confidence	.04 ± .08
Unsociability	.24 ± .06
Bell Inventory	
Social maladjustment	.09 ± .07
Emotional maladjustment	.12 ± .07

relations with scores on the first two were found to approach zero, and will not be cited here. Individuals scoring high on the social scale are said by the manual to "tend to be submissive and retiring in social contacts," while those scoring high on emotional adjustment "tend to be unstable emotionally."

In the present study, Bell Emotional Adjustment was found to correlate .60 ± .04 with Bernreuter Neuroticism, indicating a substantial agreement between the two. Bell Social Adjustment, on the contrary, correlated −.05 ± .06 with the Bernreuter Sociability scale, and only .36 ± .05 and .38 ± .05 with Bernreuter Neuroticism and Confidence, respectively. That social adjustment as measured by Bell should have nothing in common with sociability on the Bernreuter-Flanagan scale is somewhat surprising.

RELATIONSHIPS WITH INTELLIGENCE

Coefficients of correlation between students' ratings as speakers and their intelligence or scholastic aptitude by the American Council Examination proved low but statistically significant (over four P.E.) for all characteristics save vivacity and quality of voice. As shown in Table 3, the abler students were in general the better speakers. Scores on the adjustment inventories, however, yielded only one correlation with intelligence which approaches significance. According to Bernreuter FS–2, unsociability goes with scholastic aptitude to the extent of .24.

SPEAKING CHARACTERISTICS AND ADJUSTMENT SCORES

Correlations between students' characteristics as speakers and scores on the two adjustment inventories are shown in Tables 4 and 5. Traits yielding coefficients amounting to three or more times their probable error are indicated by a single asterisk; those with correlations exceeding four P.E., by double asterisks. It is obvious that the majority show no relationship greater than might occur by chance. Even when corrected for attenuation, the highest r is only .36, and this is

TABLE 4

Correlations Between Characteristics Revealed as Speakers and Scores on the Bernreuter Personality Inventory

Traits compared	N	r	Partial r with intelligence constant	Correlation corrected for attenuation
B1–N. *Neuroticism and*				
Vivacity	127	.09 ± .06	.09	.11
Poise	127	−.06 ± .06	−.06	−.08
Voice	127	.04 ± .06	.04	.05
Expression	127	.09 ± .06	.10	.11
Fluency	127	−.07 ± .06	−.07	−.09
Articulation	127	−.05 ± .06	−.05	−.06
General effectiveness	127	.02 ± .06	.02	.02
Mean		.01	.01	.01
F1–C. *Lack of confidence and*				
Vivacity	114	.07 ± .06	.06	.08
Poise	114	.08 ± .06	.07	.10
Voice	114	.06 ± .06	.05	.07
Expression	114	.06 ± .06	.05	.07
Fluency *	114	.22 ± .06	.22	.27
Articulation	114	.11 ± .06	.10	.13
General effectiveness	114	.13 ± .06	.12	.14
Mean		.10	.10	.12
F2–S. *Unsociability and*				
Vivacity *	114	.18 ± .06	.14	.21
Poise *	114	.20 ± .06	.13	.26
Voice *	114	.22 ± .06	.19	.26
Expression **	114	.29 ± .06	.24	.36
Fluency	114	.16 ± .06	.08	.20
Articulation *	114	.18 ± .06	.13	.21
General effectiveness *	114	.22 ± .06	.16	.24
Mean		.21	.15	.25

partly attributable to the common factor, intelligence. To particularize:

1. Tendency to neuroticism (B1–N) yields coefficients approaching zero with all speech ratings. The average is .01, whether or not intelligence is held constant or correction made

TABLE 5

Correlations Between Characteristics Revealed as Speakers and Scores on the Bell Adjustment Inventory

Traits compared	N	r	Partial r with intelligence constant	Correlation corrected for attenuation
B3. *Social maladjustment and*				
Vivacity	152	−.08 ± .05	−.10	−.09
Poise **	152	−.20 ± .05	−.25	−.26
Voice	152	−.11 ± .05	−.13	−.13
Expression	152	−.09 ± .06	−.13	−.11
Fluency	152	−.14 ± .05	−.19	−.17
Articulation	152	−.08 ± .06	−.11	−.09
General effectiveness	152	−.11 ± .06	−.15	−.12
Mean		−.12	−.15	−.14
B4. *Emotional maladjustment and*				
Vivacity **	152	.20 ± .05	.18	.24
Poise	152	.00 ± .06	.04	.01
Voice	152	.03 ± .06	.01	.04
Expression	152	.13 ± .05	.10	.16
Fluency	152	.06 ± .06	.02	.08
Articulation *	152	.16 ± .05	.14	.19
General effectiveness	152	.12 ± .05	.09	.14
Mean		.10	.07	.12

for attenuation. Since the Bernreuter scale for introversion gives substantially the same results at B1–N, it is safe to infer that introversion or the opposite likewise finds no reflection in the speaking traits rated.

2. Lack of confidence tends to go with better than average speech; but only with reference to fluency, or evenness of delivery, does the correlation exceed three P.E.

3. There is a small but consistent tendency for the less sociable to rate well as speakers. The *r*'s average .21 before, and .25 after correcting for attenuation. This relationship, however, depends largely on the fact that the more intelligent

are more independent and less gregarious according to the F2–S scale (see Table 3). With intelligence held constant, the mean coefficient dwindles to .15.

4. In contrast to the foregoing, those scoring socially maladjusted on the Bell Inventory are at some disadvantage as speakers, though poise is the only trait to show a significant correlation. When intelligence is held constant, the coefficients rise slightly.

5. Emotional maladjustment according to the Bell scale is likewise associated with better than average speed ratings, but the correlations are insignificant with one exception. Nervous and emotional tension appears to make for greater vivacity ($r = .24$).

ALTERNATIVE INTERPRETATIONS

A possible explanation for the comparatively negligible correlations discovered might be lack of validity in the appraisal of speech. Analysis showed that the ratings assigned by instructors A and B gave the most nearly normal distribution and least indication of halo effect. Accordingly, the ratings of these two men were combined and correlations computed with each of the other variables. The resulting coefficients differed little from those for the total group in the case of Bernreuter scores, but averaged .04 higher with the Bell Inventory. The only statistically significant relationships discovered were the same three as in Tables 4 and 5. Had all the judges been as discriminating as instructors A and B, therefore, the correlations obtained would probably have run several points higher, but the conclusions reached would not be altered materially.

A further explanation might be found in the selective factors which lead some students to elect courses in speech and others to avoid these in favor of written composition. To eliminate the possible distorting influence of such selection, it would be necessary to obtain comparable data from a wholly unselected sampling of entering freshmen.

SUMMARY

Previous investigations have shown that persons hearing the voices of individuals reading aloud show marked agreement in their judgments as to the characteristics of the readers. These judgments, however, bear little relation to the speakers' self-appraisals on personality inventories. Stagner found this true even when the readers were visible to the audience.

The present study concerns the relationship between outstanding speaking traits and personality adjustments as revealed by tests of the "inventory" type. Qualified speech instructors served as judges, and one hundred fifty six college students as subjects. Each boy or girl spoke briefly and more or less extemporaneously on two occasions and on two different topics. Despite the greater opportunity for self-revelation afforded by a speaking situation, the findings prove as negative as in experiments based on reading.

1. The individual's characteristics as a speaker are associated with abstract intelligence or scholastic aptitude much more than with neuroticism, self-confidence, sociability, or social and emotional adjustment, as measured by the Bernreuter and Bell Inventories. This holds true even after correcting for attenuation.

2. Relationships between the speaking qualities considered and the more generalized personality "traits" purportedly measured are extremely small. Only three correlations proved statistically significant. These would indicate that the less gregariously inclined speak with better expression ($r = .36$), that socially maladjusted speakers tend to be more self-conscious and less at ease ($r = .26$), and that the emotionally maladjusted display more vivacity ($r = .24$).

3. Those students making comparatively unfavorable showings on the personality inventories were on the average among the more intelligent and better speakers.

4. A curious incidental observation is that scores on the Bernreuter-Flanagan scale for sociability have nothing in common with social adjustment as measured by the Bell Inventory ($r = -.05 \pm .06$).

THE LIMIT OF USEFULNESS OF ACCIDENT RATE AS A MEASURE OF ACCIDENT PRONENESS

PERCY W. COBB

Highway Research Board, National Research Council, Washington, D. C.

IN estimating an ability by means of tests, it is necessary first to define that ability. Then a significant number of persons of known ability, as so defined, must be subjected to the tests. A statistical comparison between the test results and the known abilities will enable us to determine to what extent the tests may be truly said to measure the ability. In appraising the usefulness of automobile driver tests by this method, the definition of ability, (or rather of inability), has reduced itself to the matter of accidents, and the measure used has been accidents per year of record. There are questions which arise in this connection; for example, whether such a datum should not be weighted in some way, since a record of one accident in one year can not possibly be as significant as another of ten accidents in ten years, although the empirical rate is identical in the two cases. But it is not proposed to go into a general discussion of this aspect of the subject at present.

It has been said that there is no such thing as luck. Such a statement may have the pragmatic value of inducing the hearer to examine his own habits and conduct to good advantage, but it is nevertheless untrue. The success of a card-player on a particular occasion will without doubt depend upon his skill in the game. But it is equally certain that it will depend also upon the cards that are dealt him. The skill depends upon the person and is, conceivably at least, to be arrived at from the consideration of his biography and an appropriate examination of the person himself. On the other

hand, the cards dealt him, presupposing an honest shuffle and deal, are in no conceivable way to be predicted from an examination of the person himself. The deal is governed by causal antecedents unrelated to the person.

There is also a large element of luck in the fall of accidents, from which it follows that in a sample of cases of equal liability, the accidents do fall unequally, and even more unequally than the nearest possible approximation to uniformity. This is an hypothesis from which a form of distribution has been deduced known as Poisson's distribution (1). The mathematical statement of it gives, as the expected number of persons who will incur 0, 1, 2, etc. accidents, the corresponding terms of the series:

$$N e^{-m} (1, m, m^2/2, m^3/3!, m^4/4!, \text{etc. . . .}) \qquad (1)$$

in which N is the number of persons considered and m is the mean, or the quotient of the number of accidents divided by the number of persons. In another form the same thing may be stated in this way: the expected number of cases in which p accidents occur will be:

$$N e^{-m} m^p/p! \qquad (2)$$

Since this is the form of distribution consequent to the hypothesis of equal liability in all cases, we may say that when such a distribution is found, it is useless to look for differences in liability among the cases. The differences in the empirical data are fully accounted for by the unrelated operation of "luck."

As a matter of fact, none of the collections of automobile accident data which have come to the writer's attention could be shown to follow this law (2, 18–28, Tables 5–13). Two sets of factors in the driver's success in keeping out of accidents are therefore operative. There appears to be a group dependent upon the driver; and there are, collectively, the unrelated factors. The actual distribution of accidents is reasonably to be interpreted in this way: the whole population is to be thought of as the aggregate of many partial popula-

tions, each composed of members of equal liability, and each partial population having a different liability from that of any other. To state this hypothesis is one thing. To resolve a population of drivers into such groups, even numerically and without identification, is another matter.

The liability to accident of a given individual is an hypothetical value. It may depend upon at least two sets of factors. One of these sets may be supposed to be related to such of his practices as will of themselves increase his exposure to accident, as for example, the number of miles per year he customarily drives, and whether these miles are driven under more hazardous, or less hazardous, conditions. The other set relates to his personal aptitude as a driver, his bodily competence, his alertness, his judgment in reading hazards in advance, etc. These two sets, taken together, may be termed the related factors, and are those which can conceivably be determined from his biography and by an examination. The other factors are, collectively, unrelated to him, and fall into the category which we may call luck.

The point of this discussion is this. If we assume the impossible and suppose that we are possessed of a perfect driver testing system, which must of course take into account all of the factors contributing to the subject's liability, the best we could conceivably expect would be that a group selected by such a system from a random sample of the population would fall into a Poissonian distribution, since the testing system has enabled us to estimate, and so to equalize as to their combined effect, the personal and related factors within the group.

This leads to a simple way of appraising the largest coefficient of correlation that could possibly be obtained within an unselected population, between the prediction of an ideally perfect testing system and the empirical record of accidents. The limit so determined can not conceivably be exceeded, and can not be attained in practice. The method of computing it will be best made clear by the use of an actual record.

A sample of 29,531 drivers in the state of Connecticut, licensed and driving during the whole period of six years,

1931–1936, showed the distribution of recorded accidents given in Table 1 (2, 19, Table 5). Here the accidents are 7082 in number, and the rate, or mean accidents per person, is 0.2398.

It is a mathematical property of the Poissonian distribution that the variance (the square of the standard deviation) is equal to the mean (or rate). If this sample is now separated into a number of classes, each of equal related liability within itself, these classes will constitute true Poissonian distributions and, within each one, the mean will be equal to the variance. The means of these classes will, of course, be on the regression line of the table so formed. The weighted mean of the means will be the mean of the whole sample. And the weighted mean of the variances within the classes will also, since the mean and the variance are equal for the class, be equal to the mean of the whole sample. *But this weighted mean variance will also be the variance from the regression-line,* which added to the variance due to regression is identical with the variance of the whole sample.

Now the variance of the sample in Table 1 is 0.3056. The mean of the sample is 0.2398. The variance due to regression is then 0.3056–0.2398, or 0.0658, and R^2 is therefore 0.0658/0.3056, or 0.215, whence $R = 0.464$ (3).

TABLE 1

Accidents per driver	Number of drivers	Accidents sustained	
x	y	xy	x^2y
0	23881	0	0
1	4503	4503	4503
2	936	1872	3744
3	160	480	1440
4	33	132	528
5	14	70	350
6	3	18	108
7	1	7	49
Total	29531	7082	10722
		Mean = 0.2398	0.3631
			− 0.0575
			Variance = 0.3056

Thus, if we should apply the perfect testing system to this population, this is the highest correlation we could possibly get. And this will be a limit effective in two ways. It will limit the reliability of the experimental validation of the tests; and, if the tests are to be applied as a part of the official routine to all applicants for drivers' licenses, it will correspondingly limit the usefulness of the tests in predicting the subsequent record of the candidate.

Certain warnings should be given here. The value for R deduced above, applies to the sample cited and to the period of observation represented by that sample. It is a sample obtained by drawing every tenth card from a file of some 400,000 names. A substantial number of the names so drawn were rejected for the reason that the person had not been driving through all of the six years of record, but there was no selection whatever on the basis of accidents sustained.

It can be shown that with a longer period of observation of the same population, a higher value of R may be expected. The mean of the whole distribution will increase in proportion to the period of observation, while the variance of the means of the classes will increase as the square of the period. From this, we may deduce that for 12 years observation, the mean (rate) for these drivers may be expected to be 0.4796, and its variance $0.4796 + 0.2632$, or 0.7428, which yields:

$$R^2 = 0.2632/0.7428 = 0.3543, \text{ and } R = 0.595.$$

Again, if the subjects are selected on the basis of accidents in any way, the relationship may be changed, for by no possibility, to illustrate, may a group of persons having one accident each in a given period be distributed according to Poisson's law, and one postulate of this method will be violated. This is not to raise an objection to such selection of subjects when the purpose of the work is the validation of the proposed tests. As a matter of fact, this may be done to advantage by calling notorious accident-makers to undergo the tests, alongside the volunteers without bad records. But

I am not prepared to say that, with such selection, the relations discussed above will prove to be valid.

The foregoing analysis is in substance part of a research report prepared by the author for the Highway Research Board and the U. S. Bureau of Public Roads, and is published with the permission of both agencies.

NOTES AND REFERENCES

1. The Poisson distribution is described and illustrated by specimen data in some of the textbooks on statistics, as: Waugh, Albert E., *Elements of Statistical Method,* McGraw-Hill Book Company, Inc., New York, 1938, 109ff. A more detailed discussion of its properties is found in: Yule, G. Udny and Kendall, M. G., *An Introduction to the Theory of Statistics,* Charles Griffin and Company, Ltd., London, 1937, 187–191.
2. U. S. Bureau of Public Roads, *Motor Vehicle Traffic Conditions in the United States: The Accident-prone Driver.* (House Document No. 462, Part 6, 75th Congress, Third Session.) Washington: U. S. Government Printing Office, 1938, Pp. x + 52.
3. The R discussed here is related to Charlier's "coefficient of disturbancy," but is different from it. Charlier's coefficient is the quotient of the standard deviation of the liabilities of the members of the population divided by their mean liability. The two are related to the empirical distribution as follows:

$$\text{Charlier's coefficient, } \rho = \sqrt{\frac{\sigma^2 - m}{m}}; \quad R = \sqrt{\frac{\sigma^2 - m}{\sigma}}$$

in which m is the mean (accidents per person) and σ is the standard deviation of the distribution.

Cf. Newbold, E. M., *A Contribution to the Study of the Human Factor in the Causation of Accidents.* Industrial Fatigue Research Board, Report No. 34, H. M. Stationery Office, London, 1926, 23–24. (Pp. v + 70.)

THE SOCIAL BASIS OF MENTAL DIAGNOSIS*

EDGAR A. DOLL

The Training School at Vineland, New Jersey

THE text for this homily is found in the following words: "It is only in the heart of society that man can attain the preeminent position which is his natural destiny." With these words Itard closes the opening sentence of his first report on the Wild Boy of Aveyron.[1] In this statement he epitomizes the social philosophy and guiding purpose of the clinical experiment from which we today trace so much of the modern movement for the care and treatment of the mentally handicapped.

The present paper is constructed on Itard's thesis with reference to the role of mental diagnosis in assisting man to attain his social destiny. For those who practise mental diagnosis are concerned to assist men as individuals to accomplish that type and degree of social adjustment which affords the most expressive fruition of their native endowments.

We may begin, then, with the premise that the immediate occasion for any mental diagnosis is always some social circumstance, and that the ultimate purpose of such a diagnosis is directed toward the solution of some individual social problem. This observation is so obvious as not to receive the attention which it merits as determining both the point of departure and the point of arrival in casework. Too frequently the clinical examination begins too far down the road from the original problem and stops too far short of its real solution.

* Read before the Section on Clinical Psychology of the International Council for Exceptional Children, Pittsburgh, February 23, 1940.

[1] Itard, Jean-Marc-Gaspard. The Wild Boy of Aveyron (Trans. by George and Muriel Humphrey). New York: The Century Company, 1932. 104 pp.

This is not to say that all mental casework is concerned with deficiency, defect, maladjustment, or individual failure in self-expression. Increasingly today the clinical psychologist is faced with the more positive problems of prophylaxis and guidance designed to foster normal aptitudes rather than to overcome deficiencies. On the negative side the clinician is concerned with disability and disorder of behavior as handicaps to be corrected or compensated. On the positive side we are concerned with the optimum exploitation of normal talent, normal conduct, and normal aptitudes. But in either case the basic origin as well as the ultimate goal of mental casework is one of social concern for individual welfare. We deal with man in relation to his environment. In the total disposition of a given case they are inseparable, however specifically we may analyze the component elements or their interaction.

As a second premise we may observe that mental diagnosis as generally practised today by clinical psychologists reveals an exaggerated reliance on mental tests. We may further note that this false confidence is aggravated by limited discretion in the selection of such tests as to scope and variety. Nor can one escape the conclusion that clinical-psychological mental diagnosis is too often unsound because of limited sophistication in the interpretation of such mental test data as are obtained. The tests employed are too frequently selected without specific regard to the original problem and purpose of the examination, and reveal a random or routine choice of tests for their own sake rather than for their peculiar fitness in the given situation. There is, in short, too little analysis of the clinical problem as a whole in terms of social origin and social outcome, with the result that the individual may undergo examination by devices which resemble the shotgun prescriptions of proprietary medicines.

If on the contrary the clinician begins with the fundamental social problem and proceeds from that base through the maze of secondary and more remote problems to the clinical analysis of the individual in relation to his developmental

history, he will then be in a more sound position to determine the nature and course of the mental examination most likely to clarify those problems for purposes of treatment. To this end the clinician needs to employ some over-all appraisal of the individual which stresses no single aspect of his personality but rather assays the whole. From this point of vantage he may then proceed with a more detailed analysis of the relevant variables which compose the aggregate. And with such a method the clinician is able more successfully to evaluate the success of treatment.

As a third premise we may then state that mental diagnosis is well-advised when directed toward an evaluation of the individual as a whole in his social milieu. This bio-social emphasis now receives more attention in principle than in practise for the simple reason that no satisfactory over-all method of individual appraisal has been available. Consequently, in spite of the acknowledged precept, the clinician finds it impracticable to study the entire organism except by way of its parts. He therefore is compelled to dissect the individual for detailed analysis and then as best he can reassemble him by some new synthesis of elements.

It is this very dilemma which we sought in some measure to solve by constructing the Vineland Social Maturity Scale.[2] This scale is founded on the assumption that nothing that the individual is or does is significant except in terms of its ultimate social import. There are, to be sure, exceptions from this principle with reference to the inner soul or private life of the individual, but even these are sooner or later reflected in some social impact. This principle of social competence serves as a point of reference for evaluating all detailed aspects of individuality since each of these is ultimately cap-

[2] DOLL, EDGAR A. The Vineland Social Maturity Scale: Revised Condensed Manual of Directions. Publication of The Training School at Vineland, New Jersey, Department of Research. Series 1936, No. 3, April, 1936. Pp. 30.

An annotated bibliography of investigations and experience with this scale is now in press, *Journal of Consulting Psychology*.

italized for some social end. Thus the individual's appearance, his health, his racial or national derivation, his sensory acuity, intelligence, motor coordination, skills, attainments, personality, in short all his attributes, are ultimately significant in terms of their contribution to his personal-social welfare.

From this it follows that the clinician concerned with mental diagnosis must go beyond the mere mentality of the individual and include his total individuality. All these aspects of the unique personality are reflected in the total diagnosis as a clinical picture, syndrome or symptom-complex.

As a fourth premise, it is obvious that every individual grows in various respects as his age increases. This growth is initially determined from the potentials inherent in the fertilized cell, the final fruition of which is modified by environmental circumstances. As the organism grows in structure it develops in function. The individual also matures in living as he grows in size and develops integrated organization. In this process of maturation the organism profits from environmental stimulation and opportunity through the spontaneous capitalization of experience as well as through formal instruction.

Social competence may therefore be considered in terms of its maturation at successive life ages or progressive developmental stages. An immediate problem is to determine the nature and course of this maturation, and in the specific case to measure it in terms of the over-all concept of individual adjustment. This social aspect of growth and development may be conceived in terms of individual responsibility (self-sufficiency) and group responsibility (social participation). That is, each person as he matures acquires increasing degrees of personal independence and exercises an increased sharing in the social order as a whole.

In constructing a scale for the measurement of social competence in terms of maturation we may employ two different approaches. The first consists of an inventory of those relatively universal forms of social expression which are charac-

teristic of given age periods as successive stages of development. Or we may employ the alternative of analyzing social behavior into certain gross categories of behavior and delineate the evolution of successive degrees of aptitude within these categories. In constructing the Vineland scale both methods were used. It is further necessary to designate specific modes of behavioral expression and to define these in such a way as to afford distinctive stages of their progression.

In formulating these item performances it proved desirable to employ only those details of behavior which are relatively universal in a standard environment so that departures from the norm might be interpreted either in terms of individual aptitude or in terms of environmental opportunity and convention. To this end it was necessary to select items which are relatively free from the influence of intelligence as such, as well as of those other personal traits and environmental conditions which might affect a given item too heavily in the direction of specific variables of less universal distribution. The categories selected as reasonably representative of total behavior include such universal performances as self-help, locomotion, occupation, socialization, communication, and self-direction. All individuals mature in each of these respects in any environment. The specific mode and tempo of expression in these respects may be peculiar to a given environment, but are relatively universal within that environment with due regard for individual differences. Indeed the influence of the environment may itself be evaluated by observing those modifications of the scale which are necessary in alternative environments.

This scale has been released in experimental form for professional experimentation. It has not yet been published as an instrument for general use justified by ample details of standardization and validation. During the experimental period, however, many studies have been made with different types of individuals designed to explore the technical advantages and weaknesses of the method and to provide an ample experimental foundation for ultimate release of the scale as a

practical instrument. In the meantime, many clinics and individual clinicians are finding the scale useful at the same time that they are evaluating its limitations. Much subsequent work will be necessary before the scale can be considered adequately standardized or reliably proved. Yet those employing the scale in its present form give many assurances regarding its scientific value and practical usefulness.

This scale affords an immediate means for determining the present, and under some circumstances even the past, social competence of the individual in terms of individual maturation. Its use is most apparent as affording an initial appraisal of the individual on the basis of his total traits and their capitalization in the environmental situation. Because of the composite capitalization of abilities reflected in the items of the scale the examiner is enabled to obtain a more comprehensive understanding of the person under examination than is otherwise practicable. Questions of health, special handicap, special aptitude, general ability, educational attainment, occupational proficiency, and the like, are broadly revealed in a manner which affords direct suggestions for specific clinical elaboration of detail. The scale also affords a means for subsequently interpreting these details in terms of a total criterion. It also affords a means of measuring the improvement of the individual as an outcome of treatment recommended as a result of the total mental diagnosis.

The argument may be made specific by an illustration. Consider the problem of mental deficiency, regarding which there is so much confusion today because of the ill-advised confidence in mental tests, or more specifically, in various IQ's. Mental deficiency has traditionally been considered a social problem. If the feeble-minded were capable of managing their affairs with ordinary discretion, or if they were capable of sustaining themselves in society without the aid of others, our interest in this condition would be only academic. As a matter of fact, the problem of mental deficiency is one of vital social concern because of the social inadequacy of the feeble-minded. They come to our attention because

of this inadequacy and our mental diagnoses are designed to indicate those modes of disposition which will best conserve the social interests of the individual, the family, and society. Feeble-mindedness is initially defined on the basis of the essential social incompetence of the individual so affected. This incompetence is variously expressed at different life ages in terms of behavior, adjustment, profit from instruction, occupational proficiency, and social-economic independence. These basic concepts of feeble-mindedness have not previously been very clearly defined nor adequately measured. This difficulty of clearly expressing the social incompetence of the feeble-minded in standard modes and degrees of behavior has led to a current disregard for the basic significance of social competence as the first consideration in the authoritative recognition of this condition.

Binet and others before and after him made clear that the social incompetence of the feeble-minded was a consequence of their mental incompetence, and the success of Binet's scale for measuring intelligence led to the gradual, and now almost total, relegation of the social criterion to the background of diagnosis. Binet and others demonstrated that the socially incompetent who are feeble-minded are also mentally incompetent. But neither Binet nor any one since has demonstrated that the mentally incompetent are inevitably socially incompetent, or that all of the socially incompetent are feeble-minded. In short, to avoid these logical errors of the obverse the concept of feeble-mindedness requires not one but three criteria, namely, social incompetence, mental incompetence, and arrested development. The socially incompetent who are not feeble-minded include many insane, epileptic, ne'er-do-wells, physically handicapped, and so on, among whom intelligence may or may not be subnormal, and among whom intelligence if subnormal may be due to deterioration or dysfunction, or to limited expression rather than to genuine developmental deficiency.

Hence it is not permissible to diagnose mental deficiency in terms of intellectual retardation alone. Before a person can

be considered feeble-minded his low intelligence must be shown to have been arrested in development and it must also be shown that the low intelligence is accompanied by submarginal social competence. Every psychologist knows, or should know, that many of those with low IQ as measured by Stanford-Binet are not necessarily of correspondingly low degrees of intelligence as measured by other tests. Many of these individuals are above the limits of social competence which may properly be designated as feeble-mindedness or mental deficiency. Hence it becomes obvious that the determination of social competence is a *sine qua non* of the diagnosis of mental deficiency, and if this principle were accepted in practise it would go far toward obviating the confusion now prevailing in both the scientific and the public press that individuals may be altered from genius to deficiency or from deficiency to genius by environmental circumstances. The current turmoil on the relative influence of nature and nurture is much more clear when confined to mental test data than when expressed in designations which are clinically conceived as diagnostic symptom-complexes. This confusion can readily be clarified by employing a measurement criterion for the social factor in mental diagnosis just as it can be clarified by using batteries of mental tests as opposed to single scales.

The same issues are reflected only in different manner and degree as the illustrations are drawn from the fields of general and special talent, delinquency, behavior problems, special disabilities, or social adjustment and maladjustment of any type. We may safely stand by the position that no mental diagnosis is complete which does not begin with a sound estimate of social competence or which does not end with the prediction or confirmation of social competence following prognosis or treatment.

This argument applies with even greater force in the case of those exceptional children and adults who are affected by handicaps of vision, hearing, speech, or motor coordination, or in short any of the groups encompassed within the interests of the International Council for Exceptional Children.

In support of the observations that current mental diagnosis frequently employs an ill-advised selection of test procedures, places too much reliance on mental tests, and is inclined to unsophisticated interpretation of such data, we may now more logically reassert that every case study originates in some social situation. This is commonly referred to as the immediate problem or the "complaint." Sound casework requires that this immediate problem be formulated as explicitly as possible, and that it be verified in fact as may be practicable by objective procedures before proceeding with the clinical analysis of the problem. The delineation and verification of the initial problem requires a review of the individual's history in all aspects and this review is most serviceable when obtained prior to the use of test procedures. One of the most serious errors in current clinical-psychological practise is to proceed with test devices long before the immediate problem is specifically assured and before the antecedent history of the individual has been adequately explored. We may suggest with confidence from our own experience that the use of the Vineland Social Maturity Scale serves greatly to illuminate the immediate problem and sheds critical light on its antecedent origins as it also indicates the most profitable lines of clinical exploration for purposes of diagnosis and treatment. The only sound position that can be taken by any competent clinician is to hold fire on test procedures until these preconditions of a satisfactory mental diagnosis are well established. Only by this approach can the clinician decide which tests are relevant in a given case and how many or of what variety. Only under these circumstances can he economically and quickly determine the direction of test analysis. And only from this base can he properly interpret the data derived from applying clinical-psychological test procedures with assurance of a harmonious coordination of these data with the entire clinical picture. And we might add that only in this way can the psychological clinician also anticipate the desirability or necessity of referring a case for consultation with other specialists outside his own field.

We may conclude as we began with a reiteration of Itard's thesis, namely, that "It is only in the heart of society that man can attain the preeminent position which is his natural destiny." The role of the clinical psychologist in facilitating the attainment of that destiny consists in a harmonious evaluation of all the circumstances, both in the individual and in the environment, which are thereto related. This requires a final integration of mental diagnosis based on evaluation of the individual within his social circumstances. It requires a clear understanding of the immediate reasons for a given examination and its ultimate purpose. This point of view particularly demands a clear understanding of the social setting within which the problem has developed, that is, the immediate social circumstances which precipitate the examination. And it particularly involves clear concepts as to the social outcome of the mental diagnosis to be accomplished through treatment as indicated by the total clinical analysis.

BOLDNESS AS A FACTOR IN TYPE-DESIGN AND TYPOGRAPHY

MATTHEW LUCKIESH AND FRANK K. MOSS

Lighting Research Laboratory, General Electric Company, Cleveland

AMONG the controllable physical factors which facilitate ease in reading, the level of illumination and the size and boldness of type have been shown by previous researches to be the most effective (1). Obviously, these are the factors which primarily determine the intensity of the visual sensation—or the visibility of the object of regard. However, it is not implied that the configuration or design of the individual characters is inconsequential, but rather that this factor is less important, in types commonly accepted for general use, than illumination, size and boldness. These facts have led us to emphasize the factor of boldness in our quantitative studies in typography (3).

The specific objective of the present investigation is the determination of the optimum degree of boldness for a specific family of 10-point types as indicated by the criteria of visibility (measured by the Luckiesh-Moss visibility meter) and readability (rate of involuntary blinking). The Memphis family was chosen because several degrees of boldness were available and our previous investigations of type-design or style revealed the fact that it is a representative type. The data so obtained are interpreted on the assumption that the type found to be best is to be used for body text. Since these criteria do not necessarily appraise the same characteristics of the reading-matter, it is possible that the criteria of visibility and readability will indicate different optimum degrees of boldness. This possibility introduces a secondary objective of the present research—namely, the determination of the extent to which

visibility measurements indicate the attribute of readability as the latter is influenced by boldness.

A Measure of Boldness. In determining and interpreting relationships between boldness and visibility and between boldness and readability, a quantitative measure of boldness is essential if distortions and ambiguity are to be minimized. In other words, the value of precise measurements of visibility and readability is decreased if these data are related with mere qualitative descriptions of boldness such as "light," "medium," etc. Therefore, we have devised a precise photometric method for appraising the relative boldness of types of the same size and amount of leading. A description of this method follows.

We have assumed that the degrees of boldness of a series of types of the same size, alphabet length and leading, are proportional, respectively, to the areas of the inked surfaces. The latter are, in turn, proportional to the differences in the light diffusely reflected by the printed and unprinted page. Such data are readily obtainable with the Taylor Reflectometer (5) —an instrument for measuring the reflection-factor of any surface. For example, a page printed in the boldest type will reflect the least amount of light and hence will possess the lowest reflection-factor. Therefore, the difference between the diffuse reflection-factors of the unprinted paper and the printed page will be maximal in such a case. This difference is assumed as a measure of the boldness of the printed matter. Obviously the measurements of reflection-factor of the printed page should be made only for areas of full body text and under a standard condition of diffuse illumination. Several such areas should be measured and averaged.

The reflection-factors of pages printed in four weights of Memphis types with two points of leading on non-glossy white paper (reproduced in Fig. 1) are given in Table 1, as well as the reflection-factor of the so-called white paper used in preparing the typographical samples. It will be noted that the measurements made by different observers are quite consistent. These data indicate that the boldness of the members of this

Meanwhile the plough does it work year by year, the harvests are gathered, the builders build as they are told, the tradesmen work and acquire fresh devices; the knowledge of writing spreads; novel things, the horse and wheeled vehicles and iron, are introduced and become part of the permanent inheritance of mankind; the

MEMPHIS LIGHT

At a remote period in geological time there is good reason for supposing that there was land where now the Atlantic waters roll, but there is no evidence for and much against any westward extension of Europe or Asia since the Miocene period. But civilization is a matter of at most the last twenty thousand years and probably

MEMPHIS MEDIUM

Very soon the seafaring men must have realized the peculiar freedom and opportunities the ship gave them. They could get away to islands; no chief nor king could pursue a boat or ship with any certainty; every captain was a king. The seamen would find it easy to make nests upon islands and in strong positions on the main-

MEMPHIS BOLD

The trade that was going on in the ancient world before the sixth or seventh century B.C. was almost entirely a barter trade. There was little or no credit or coined money. The early empires got along without coin altogether. The ordinary standard of value with the early

MEMPHIS EXTRA BOLD

FIG. 1. Four weights of 10-point Memphis type with 2-points of leading and composed in 21-pica lines. Text from "The Outline of History" by Wells.

family of 10-point types does not increase exactly in either an arithmetic or a geometric progression. While it cannot be stated categorically that a series of types should follow an exact progression, there are ocular reasons for believing that a geometrical one would be desirable. For example, it is a well-known fact that visual sensation (visibility) increases arithmetically with geometrical increases in the intensity of the stimulus (boldness of type in this case).

TABLE 1

These data represent the average reflection-factors of the several specimens of printed matter (10-point types and 2-point leading) as determined by 30 measurements by each observer. The unprinted page was so-called white paper with non-glossy surface

Observer	Reflection-factor in per cent				
	Memphis Light	Memphis Medium	Memphis Bold	Memphis Extra Bold	Unprinted page
A	68.3	65.6	63.7	58.8	81.8
B	67.9	64.7	62.9	58.0	83.0
C	68.1	65.7	62.9	58.3	81.6
Average	68.1	65.3	63.2	58.4	82.2
Difference	14.1	16.9	19.0	23.8	
Relative boldness	100	120	135	169	

It will be noted from Table 1 that the relative boldness of Memphis Medium is 20 per cent greater than that of Memphis Light. If this percentage differential is assumed for the other types in a geometrical series, the relative boldness should be 100, 120, 144 and 173, respectively. Therefore, it is of interest to observe that an approximate geometrical series (with 20 per cent differentials) would be obtained by increasing the relative boldness of Memphis Bold from its present value of 135 to a value of 144. This example illustrates the significance of exact visual measurements in the design of a family of types of various degrees of boldness. It is also possible that such a basis for design might result in minimizing the number of

faces in a given series of types of the same family. For example, in considering the use of these types only for body text it might be desirable to make Memphis Bold bolder and eliminate Memphis Extra Bold.

The Visibility of Memphis Types. The relative visibilities of the four weights of 10-point Memphis types, as determined with the Luckiesh-Moss visibility meter (1), are presented in Table 2. The data pertaining to three of these types are strictly comparable since these types are similar in point-size and character count. However, the data for Memphis Extra Bold should be corrected to an "equivalent size" for comparison with the Light, Medium and Bold types since it differs in character count.

TABLE 2

Each of these values represents the arithmetic average of 10 measurements of visibility of 10-point types of four degrees of boldness under a diffuse illumination of 10 footcandles. The order of measuring the four types was reversed in taking the second series of measurements. The two series of measurements were made on different days

Observer	Memphis Light	Memphis Medium	Memphis Bold	Memphis Extra-Bold
A	2.97	3.19	3.40	4.25
	3.43	3.62	3.74	4.38
B	4.22	4.68	4.73	5.59
	3.91	4.64	4.51	5.44
C	3.49	3.97	3.98	4.77
	3.61	3.89	4.14	4.91
D	2.89	3.75	3.86	4.61
	3.31	3.93	4.15	4.25
E	3.00	3.40	3.30	3.75
	2.88	3.34	3.29	3.61
Averages				
First series	3.31	3.80	3.85	4.59
Second series	3.43	3.88	3.97	4.52
Both series	3.37	3.84	3.91	4.56
Relative visibility	100.0	114.0	116.0	135.3 (115.5)*
Relative character count	100.0	100.0	100.0	83.8

* Corrected visibility.

It will be noted from Table 2 that the relative visibilities of Memphis Light, Memphis Medium and Memphis Bold are 100, 114 and 116, respectively. These values are comparable since the average number of characters per line is the same for all three types. However, the number of Memphis Extra Bold characters per line is 83.8 per cent of that of the other three types. On this basis, the Memphis Extra Bold characters are arbitrarily computed to be 19.2 per cent greater in "width" than the characters of the other types. It was also found by exact comparator measurements of several lower-case letters, that the Extra Bold characters (without ascenders or descenders) average 10.0 per cent greater in height. The average of these two values (14.6 per cent) is arbitrarily used as a correction-factor for the Extra Bold type and results in a relative visibility of 115.5.

It is recognized that this method of "correcting" the visibility of the Memphis Extra Bold type is an approximate one. However, it will be noted from Fig. 2 that the value of relative visibility so derived is not inconsistent with the general trend of the curve relating visibility and boldness as established by the other three types. Furthermore, since the Extra Bold would scarcely even be considered for extensive text-matter, some error in correcting the visibility of this type is not a matter of practical importance. It should also be obvious that the indicated value of relative boldness, 169, would also be altered somewhat by reducing the size of the Extra Bold characters to that of the other three types. However, any possible change in the boldness-rating of this type would not alter the shapes of the curve relating boldness and visibility to any significant extent.

The outstandingly important fact revealed by the visibility-boldness relationship (Fig. 2) is *the definite indication of an optimum boldness.* In this family of types, it appears that a practical optimum in boldness is obtained with Memphis Medium since the next step in boldness produces only a slight increase in relative visibility, from 114 to 116, which is less

than two per cent. Actually, such an increment in visibility is not significant from statistical or practical viewpoints.

The superiority of 14 per cent in visibility of Memphis Medium as compared to Memphis Light may also be expressed in terms of "equivalent" type-size (1). For example, 10-point Memphis Medium is equivalent in visibility to 11.3-point Memphis Light if the latter is assumed as a base or given a 100-per cent rating in visibility. Upon the same basis, 10-point Memphis Bold is equivalent to 11.5-point Memphis Light. Thus the difference between Memphis Medium and Memphis Bold amounts to but 0.2 of a point in equivalent type-size.

In comparison with types commonly used for body text, it is probable that Memphis Medium would be considered as a "bold" type and, perhaps, as a bolder type than is usually recommended for body text. In view of the fact that Memphis 10-point types (excluding the Extra Bold) vary in visibility over a range of about 15 per cent, it is probable that other commonly used type-faces (of different degrees of boldness) would vary similarly in visibility. Furthermore, significantly higher visibilities may be obtained by increasing boldness without unduly sacrificing the individuality of many types of lighter faces. In any event, the potential enhancement in visibility appears to be an important one. This opens an important realm of research in type-design or, at least, suggests close cooperation between the type-designer and a laboratory skilled in techniques for testing the results of new attempts by the designer to produce types of highest visibility compatible with suitable readability.

The Readability of Memphis Types. Since the rate of involuntary blinking (4) has been shown to increase when conditions for critical seeing are made more unfavorable,* it is

* By: (1) decreasing the illumination, (2) increasing the duration of the task, (3) the presence of glare, (4) increasing the contrast between task and surroundings, (5) flickering illumination, (6) erroneous eyeglasses, (7) increasing extrinsic muscle effort, (8) decreasing size of type, (9) decreasing the leading, (10) use of colored backgrounds, (11) inferior printing, (12) altering type-face, etc.

reasonable to assume that this criterion is a reliable one for appraising the influence of boldness upon readability in any family of types for body-text. The rate of involuntary blinking while reading the four weights of 10-point Memphis types, as determined with 40 subjects, is summarized in Table 3. In obtaining these data, each subject read each of the four types for 5-minute periods on two different occasions with the reading matter under a diffuse illumination of 10 footcandles. In all cases, the sequence in which the types were read was reversed in the second series of measurements in order to cancel possible progressive increase in the rate of blinking as the duration of the reading increased.

TABLE 3

The actual and relative number of blinks occurring during 5 minutes of continuous reading of 10-point type printed on dull white paper. It will be noted that the data have been averaged arithmetically and geometrically. All of the data from the 40 subjects are included in these averages. Systematic differences in blink-rate among the subjects were elminated in computing the probable errors of the averages

	Memphis Light	Memphis Medium	Memphis Bold	Memphis Extra Bold
Arithmetic mean				
Number of blinks	27.2	24.6	26.1	27.6
Relative rate of blinking	100.0	90.4	96.0	101.5
Geometric mean				
Number of blinks	21.0	18.5	19.8	20.8
Relative rate of blinking	100.0	88.1	94.3	99.1
Probable error	.....	1.9	2.6	3.2

It will be noted from Table 3 that about the same relative rates of blinking were obtained by both methods of averaging the experimental data. This is an indication of the reliability of these data since a wide variation in these averages would indicate that the individual data were correspondingly erratic. It will also be noted that the differences in the rate of blinking while reading these types are quite significant when compared to their corresponding probable errors.

For convenience in comparing the results obtained by the criteria of visibility and rate of blinking, the relationship between rate of blinking and boldness has been inverted and designated as the relationship between *readability* and *boldness*. This relationship is shown in Fig. 2. Thus for the first time in typography, the visual characteristics of a family of types are presented graphically and quantitatively from the

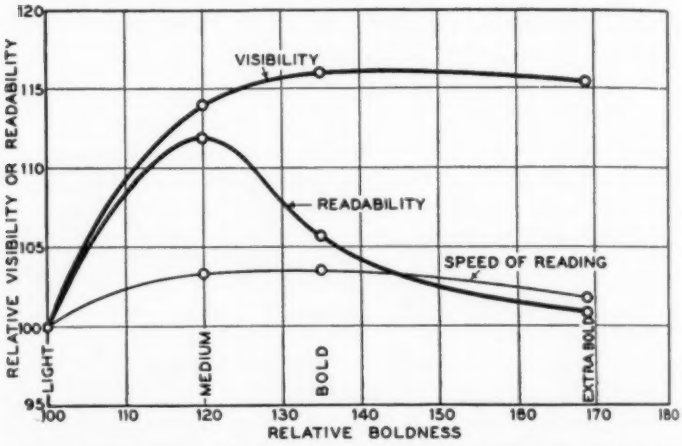


FIG. 2. The curves of *visibility* and *speed of reading* are plotted from the data presented in Tables 2 and 5, respectively. The relationship between boldness and rate of blinking (geometrical means) is inverted in this figure and designated as the relationship between *boldness* and *readability*.

comprehensive viewpoints of both visibility and readability. It will be noted from Fig. 2 that:

1. The optimum degree of boldness, as indicated by the criterion of readability, is obtained with Memphis Medium;
2. The optimum degree of boldness is much more sharply defined by the criterion of readability than by that of visibility;
3. A somewhat bolder type is indicated by the criterion of visibility than by the criterion of readability.

These data reveal the necessity of distinguishing between the attributes of visibility and readability of the printed page when the factor of boldness is to be appraised. It is also evident from Fig. 2 that the optimum degree of boldness is determined by factors which are highly critical in their influences upon the processes of reading. In view of these complex relationships, it seems obvious that introspective methods for appraising the influence of boldness can possess little intrinsic significance and reliability. It should also be obvious that utilitarian aspects of type and typography are not appraisable by the esthetic sense.

Although the curves relating visibility and boldness, and readability and boldness, are dissimilar, it will be noted that they are in agreement over the range in boldness between Memphis Light and Memphis Medium. Thus it appears that the criterion of visibility is appropriate for appraising the influence of boldness on ease of reading for relative boldnesses within this range. This fact is important since visibility measurements appear to offer a practical basis for quantitative studies of the many details of type-design. Thus from the practical viewpoint of type-design, an important contribution of readability criteria lies in the confirmation of the appropriateness of visibility measurements under certain definite conditions and limitations. In the absence of additional data, these conlusions are at present restricted to 10-point types set with 2 points of leading (2). Even with these limitations, the criterion of visibility may be applied in the determination of the advantages to be gained by marked increases in the boldness of a wide variety of types now used for body text.

The data of Fig. 2 also reveal that (a) the readibility of Memphis types decreases sharply as the boldness is increased from that of Memphis Medium to that of Memphis Bold; and (b) that the decrease in readability from Memphis Bold to Memphis Extra Bold is less precipitous. The latter result is probably due to the much higher visibility of the Memphis Extra Bold as compared to that of Memphis Bold. This result again emphasizes the importance of type-size; and (c) that

the Memphis Bold is superior to Memphis Light from the viewpoints of both visibility and readability. This fact further supports our conclusion that boldness is an advantage to the reader which has not been generally or generously enough recognized in type-design and typography.

Although the indicated differences in readability among the 10-point Memphis types appear to be real from a statistical viewpoint, the significance of these data would be greatly enhanced if approximately the same relative values were obtained under somewhat different experimental conditions. Accordingly, the investigation was repeated by a different experimenter under the following conditions:

1. In a different laboratory and under a level of illumination of 5 footcandles instead of 10 footcandles. Thus the environmental conditions under which the reading was done were completely altered.
2. The order of presenting the several weights of Memphis types was altered. In the first investigation, the orders were (a) Light, Medium and Bold; and (b) Bold, Medium and Light. In the second investigation the orders were (a) Medium, Light and Bold; and (b) Bold, Light and Medium. It will be noted that the positions of the type of highest readability (Memphis Medium) and of the lowest readability (Memphis Light) were transposed in the second investigation.
3. The measurements were made with 18 instead of 40 subjects, each of whom read each type on four occasions instead of twice as in the first investigation. Since the same materials were read, the rates of blinking obtained in the second investigation were those resulting during a rereading of the text-matter for the second and third times.

The relative rates of blinking in both investigations are given in Table 4.

The agreement between the results of these two investigations emphasizes, not only from a qualitative but also from a quantitative viewpoint, the inherent significance and reliability of the criterion of involuntary blinking. Specifically, the statistical possibility that this agreement is due to chance is practically

TABLE 4

The relative rates of blinking while reading three weights of Memphis 10-point type as determined in two independent investigations

	Light	Medium	Bold
Arithmetic means			
First test	100.0	90.4	96.0
Second test	100.0	88.9	94.5
Geometric means			
First test	100.0	88.1	94.3
Second test	100.0	88.8	95.0

nil. Considered in conjunction with the fact that reproducible results were also obtained in repeated investigations involving the variables of illumination and brightness of the area surrounding the printed page, it would seem that an acceptance of the reliability of the blink-rate criterion is inescapable.

Speed of Reading. In the present investigation, the subjects were instructed to read at their normal rates and the rates of reading thus obtained (Table 5 and Fig. 2) again confirm our long-standing conclusion that this criterion is a relatively insensitive and inadequate one for appraising readability. Obviously, no reliable determination of the optimum in boldness can be made upon the basis of these data. We have included this criterion in the present investigation largely for reasons of theoretical completeness. In general, our researches indicate that readability is more readily and certainly revealed by certain involuntary reactions of the reader than it is by performance or "production" over which the subject exercises voluntary control. However, the results obtained with the criterion of speed of reading are in qualitative agreement with other data as will be noted from Fig. 2.

Summary and Conclusions. Since our typographical studies are based upon quantitative appraisals of both visibility and readability, and since each investigation was prosecuted to the extent necessary to obtain statistically reliable data it is

TABLE 5

These data give the average number of lines of print read during a period of 5 minutes of continuous normal reading of 10-point Memphis types under 10 footcandles, as determined by 40 subjects. The values given for Memphis Extra Bold have been corrected to give the same number of characters per line as other types

	Memphis Light	Memphis Medium	Memphis Bold	Memphis Extra Bold
Lines read	135.2	139.6	139.9	137.7
Relative rate	100.0	103.2	103.5	101.8

possible to present certain conclusions pertaining to the utilitarian merit of various type-faces. These conclusion are:

1. A marked enhancement in the readability of the printed page can be obtained by augmenting the boldness of many types which are now being recommended or widely used for body text.
2. An enhancement in readability is decisively less promising by means of alterations in the configuration of many commonly used type-faces than by utilizing the optimum degree of boldness.
3. The design of a type of optimum readability may now be guided with reliability and exactness by measurements of visibility within a range which is now fairly definitely known.
4. The rate of involuntary blinking while reading under controlled conditions appears to be well established as a criterion of readability.

It is emphasized again that all these conclusions are based upon the intrinsic visibility of the types studied and upon the facility with which they are *actually* read and not upon introspective appraisals of the appearance of the printed page.

We are indebted to the Mergenthaler Linotype Company, Brooklyn, for furnishing the typographical materials; and to S. K. Guth and A. A. Eastman for obtaining the laboratory data.

REFERENCES

1. LUCKIESH, MATTHEW, and MOSS, FRANK K. The science of seeing. New York: D. Van Nostrand Company, 1937.
2. ————. The effects of leading on readability. *J. Appl. Psychol.*, 1938, 22, 140.
3. ————. The visibility and readability of printed matter. *J. Appl. Psychol.*, 1939, 23, 645.
4. ————. A summary of researches involving blink-rate as a criterion of ease of seeing. *Illum. Eng.*,* 1940, 35, 19.
5. TAYLOR, A. H. The measurement of diffuse reflection-factors and a new absolute reflectometer. *J. Opt. Soc. Amer.*, 1920, 4, 9.

* Beginning with Vol. 35, the name of this journal was changed from "Transactions of the Illuminating Engineering Society" to "Illuminating Engineering."

THE CASE RECORD IN PSYCHOLOGICAL COUNSELING[1]

THEODORE R. SARBIN
University Testing Bureau, University of Minnesota

I. INTRODUCTION

FEW who have read case notes written by psychological counselors will disagree with Whitehead's statement that "the success of language in conveying information is vastly over-rated, especially in learned circles." Perhaps one of the most difficult forms of expression is that of writing meaningful case notes. In trying to record material, psychologists have variously employed three general forms: (a) the one-phrase or one-line digest, (b) attempts at photographic reproduction, and (c) selection and interpretation.

The one-phrase or one-line digest crams into a small space the counselor's[2] analysis, diagnosis and treatment. This method is usually employed where cumulative record forms or other prepared forms are the sole instruments for the recording of case data. A full hour interview might be recorded as follows: "12–1–39. Cannot make up his mind about vocation. Discussed scientific fields." Obviously this oversimplification and over-compression provides hardly any information on which to make an adequate diagnosis. If the

[1] Adapted from a paper read before the 17th Annual Meeting of the American College Personnel Association, February 21, 1940, at St. Louis, Missouri. This paper is primarily concerned with the structure of the case record in the psychological counseling of college students. The writer's experience with social work and child guidance records suggests the adoption of a similarly designed method of recording case data.

[2] The expression "counselor" or "clinical counselor" in this paper has as its referent an individual trained in clinical psychology and allied fields. See Williamson, E. G., *How to Counsel Students*, New York, McGraw-Hill Book Company, 1939.

student were subsequently to be interviewed by another counselor, he would have to repeat his entire history and complaint.

The reproductive method, on the other hand, attempts to reconstruct the whole interview from beginning to end photographically; that is to say, the counselor tries to include, usually in chronological fashion, *everything* that happens without regard for utility, significance, or relevancy. Thus the material in the case record is all-inclusive, unselected, and entirely dependent upon the memory and fluency of the counselor. This method calls to mind the attentive student whose lecture notes missed nothing, not even the professor's "Good morning."

The third method, that of selecting and interpreting elusive interview data and objective test results, requires a considerable degree of skill. One is repaid for developing this skill, however, in that the results have meaning and utility for both the counselor and the clinic. Few will deny that much of the counseling process has no relevance for the real or imagined problems of the client or patient. Establishing rapport, for example, is a necessary part of every interview, but it need not be described in detail unless it is specifically related to the complaint or problem. In constructing the case notes this material is culled out, and only that which has meaning in terms of the individual's problems is retained. Not only does the clinical counselor sift out the relevant from the irrelevant, he also interprets and gives meaning to otherwise meaningless test scores and interview data.

The first method, that of the one-phrase or one-line digest, is rapid and economical, but is to be eschewed in clinical counseling because it adds nothing to our knowledge of the patient. The second method, complete and unselected reproduction, is undesirable because of the difficulty of conveying photographic images by the hurriedly written word, and further, because of the inefficiency of recording a mass of irrelevant data. The third method, although subject to errors of judgment, is considered the most desirable in terms of the functions to be served by case notes.

II. FUNCTIONS OF CASE NOTES

Our preference for the third method is based upon a consideration of the purposes or objectives of these records of counselor-counselee relationships. Recording the events of the interview without guideposts in the form of objectives of case notes lacks significance for the counselor and for the clinic. In writing or dictating case notes, the counselor should be guided by the functions that are to be performed by his record. Case history notes, then, should be functional, fulfilling certain clearly defined objectives.

We can establish the objectives of recording counseling procedures by seeking answers to this question: what uses are made of case notes? Once we have answered this question, we shall have before us definite objectives which may serve as silent presuppositions, as conceptual sifting devices, in recording the counseling process. At least six answers may be given to our question.

First, the case record serves as a point of departure for the counselor in interviews subsequent to the initial contact. The counselor saves time for himself and for the student by reviewing the events of previous interviews before the student's re-appearance. Furthermore, he is able to identify the student by means of the case notes. It is not uncommon to find counselors who disdain case notes, treating one student for another student's problems simply because no adequate record is available of the earlier diagnostic interview.

Second, case notes serve to acquaint a second counselor with the case when the first cannot continue the treatment. In the event of a counselor's separation from the clinic, it is obviously necessary to have a record of his relationships with the student. In some cases it is necessary for one counselor to transfer a student to another counselor because of failure to maintain rapport or because of inability to cope with certain problems. Thus, case notes should serve to tell the new counselor what happened in interviews with previous counselors.

Third, understandable case notes are prerequisite to the

supervisory functions in the psychological clinic or counseling bureau. In order for the director or supervisor to know what the counselors are doing and the details of the counseling procedures, he must have some form of record available. From the examination of case notes, the director may compare the work of one counselor with another and the work of the counseling staff as a whole with the major objectives of the clinic.

Fourth, case notes are used in the training of counselors. One of the most frequently used educational techniques for acquainting interns with procedures of diagnosis and treatment is the careful analysis and study of counselor's case notes.

Fifth, case notes are often employed to correct erroneous impressions left by the student in relating the events of the interview to others. Psychological counselors are often called upon by school officials, psychiatrists, social workers, employers, and parents because of errors or inconsistencies in the student's report. Not infrequently, he may distort or misinterpret the information or advice given by the counselor. For example, an irate parent calls upon the counselor with this question: "Why did you tell my boy to quit school?" The case notes—the significant parts of which are read to the parent—tells that the counselor advised that the boy drop only one of his courses. The documentary evidence satisfies the parent. A psychiatrist suspects the student's report that "The counselor said I had a bad personality." The counselor saves his professional reputation by showing the psychiatrist the case notes which read: "I told the student that his personality traits were not like those of students in the Engineering college, that he would fit into the Arts college group much better."

The most carefully given advice is sometimes subject to misinterpretation, misquotation, and distortion. The only means whereby the counselor can protect himself from the consequences of such inaccurate reporting is adequate documentation. By employing case notes to protect himself against such contingencies, the psychologist builds and maintains his professional reputation.

Sixth, case notes are used in the facilitation of personnel research. Increasingly, case records are serving as the raw data in research. At the present time two research problems are paramount in counseling work. One is the problem of evaluation. Clearly written case notes, intelligently interpreted, serve as the first step in evaluation studies. The second research problem is the isolation and description of treatment or counseling techniques. The literature of clinical psychology is replete with discussions of analytic and diagnostic techniques, but not many treatment techniques are described in detail. Case notes may serve as an excellent vehicle for describing the treatment phase of the interview. To be sure, it is often difficult to transform a gesture or an inflection of the voice into the written word, but we must try to record these so-called intangibles if we are to make progress in psychological counseling.

To recapitulate, six major purposes of case notes are listed:

(1) Recording of the counseling process for the counselor's future use in treatment.
(2) Recording for continuation of treatment by other counselors.
(3) Supervision of counseling staff.
(4) Training of interns.
(5) Protection against misinterpretation and misquotation.
(6) Research in clinical psychology.

III. CHARACTERISTICS OF GOOD CASE NOTES

Having established six objectives of case recording, it is in order to discuss what goes into the record and what characteristics are necessary to fulfill our objectives. Before discussing the actual content of the counseling record, we first consider what characteristics good case notes should possess to carry out the previously designated functions. At least these five characteristics are essential: accuracy, brevity and conciseness, relevancy, ease of reference, and uniformity.

Accuracy: The record must provide an accurate account of the counseling process. It must reduce the inaccuracies which

would follow from mere dependence on the counselor's memory. Inaccuracies are usually the result of the lapse of time between making an observation and recording it. "Other things being equal . . . the longer the time has elapsed between your conversation and the recall of it, the less complete and the less trustworthy will be that recall."[3] At the University of Minnesota Testing Bureau we try to reduce to a minimum inaccuracies due to lapses of time by providing each counselor with a dictaphone. The moment the student leaves the consultation room, the counselor records his case notes on the dictaphone cylinder.

Brevity and Conciseness: The record should not be so long as to approach a novel in length. Service to the client or patient, not literary achievement, is our aim. We are concerned with a problem that must be solved, not with story telling. This means that minutiae and insignificanta must be omitted. The interpretations that the counselor makes will reduce the amount of material to be recorded, although they cannot substitute for concrete, factual data. The amount and kind of data will, of course, depend upon the objectives of the clinic or bureau.

It is not always necessary to write grammatically complete sentences. Many times just a phrase will carry the meaning intended, such as, "Cooperation poor on Stanford-Binet," "Performance on all ability tests in the 90's," or "Poor risk for college work." It is also considered advisable to use parenthetical remarks, which, to be sure, may interfere with an accurate chronology but which give added meaning to the content. Here is an example:

> John has low scores on the spelling sub-test of the Coop. English Test and on the reading test battery. (Later I learned that at 6 years of age his handedness was reversed—he had been a natural sinistral before. His halting speech patterns and low scores on spelling and reading tests suggest referral to Speech Clinic.)

[3] Whipple, Guy M., "The Obtaining of Information: Psychology of Observation and Report." *Psychological Bulletin*, 1918, 15, 217–248.

Relevancy: This characteristic calls for skill on the part of the counselor in determining what is relevant and what is irrelevant. If a student's problem is, let us say, definitely one of study habits and uncomplicated, it is unnecessary to include such irrelevant data as kind of magazines read or performance on an unreliable test of art appreciation. If an individual's problem is solely one of choosing a vocational goal, it is unnecessary to write about his living arrangements. A recitation of health history is unnecessary when a simple problem of social skills is the diagnosis. Careful selection of items will insure that the finished record will be a meaningful and integrated psychological document.

Ease of Reference: In the usual narrative case record adapted from social work methods, no paragraphs or identifying headings are given. Ordinarily the whole record must be read in order to discover the stated problems, the analysis and diagnosis, and the treatment. For more efficient reading, the style should be that of short paragraphs with underscored side-heads. This technique makes for rapid perusal of the record and many times will obviate the reading of the entire record when the reader is interested in finding only one particular item. Consider the usual narrative record dictated with no paragraphs or identifying side-heads. A reader interested in isolating, let us say, parental problems, would have to read through the whole record before he could discover whether such a problem was involved. If, on the other hand, a short paragraph is devoted to family or parental relationships, introduced with an appropriate side-head, then the reader would have merely to glance through the sideheads until he came to the one marked *Parental Relationships.* If the director had received a call from another agency and wanted to see if any reference had been made to the family situation, he would not be forced to digest the whole record, but could spot the paragraph immediately.

This procedure is a time-saving device in training interns in specific treatment and diagnostic techniques. If an intern seemed to be weak in knowledge or handling of family prob-

lems, he could study cases which showed diagnoses and treatments of family problems. Instead of reading every case in its entirety, however, he could isolate by means of the underscored side-heads those records which were especially pertinent to his training needs.

Uniformity: This characteristic does not mean rigid adherence to a standard form. Rather it calls for all counselors to organize their case notes and their case thinking along certain functional lines. Research especially would be facilitated if all counselors in the same clinic adopted a uniform method of case recording.

IV. A SATISFACTORY METHOD OF RECORDING CASE NOTES

To fulfill the functions outlined before, case notes should begin with the student's statement of his problem—in medical terms, the patient's "complaint." This will include the individual's reason for coming to the bureau or clinic. Here are two examples:

> Tom Jones, a junior in the College of Education, came to the Bureau with the problem: "I am not making satisfactory progress in my work." He wants the Bureau to help him discover the cause of his self-diagnosed under-achievement and to help him make the necessary adjustments.
>
> Mary Brown was referred to the Bureau by Professor X because of her persistent anxieties arising from lack of finances and from failure to make normal social adjustments.

Taking the counselor's expression of his problem or the referring agency's suggested diagnosis as the point of departure, we make an analysis to discover the cause of the complaint. Beginning with the phenotype, we search for the genotype. The analysis is made not only from test results, but from clinical data gathered in the interview and from information contained in the individual record forms. These forms usually provide such data as socio-economic status, occupation of father, work experience, and social activities, which may

or may not have a bearing upon the problem at hand. Those items that have no relation to the case as diagnosed by the counselor are omitted from the notes.

Since the interview data are the most fleeting and elusive, the counselor must dictate or write as soon as possible whatever relevant material he has collected. For ease of reference, this is given the side-head *Clinical Data.* An example follows:

> *Clinical Data:* On the individual record form the student has checked several items on the health record sheet—exhaustion, nervousness, easily fatigued, headaches, and backaches. Questioning revealed that these symptoms appear only during the school year. The headaches were especially severe when Elmer had to study. Physician's report shows "no pathology nor recommendations. . . ."

Following the recording of significant data collected in the interview, the counselor refers to the test profiles or psychographs. What data are taken from this source? A routine listing of test scores is obviously wasteful. It merely duplicates material that is already easily accessible. Suppose three tests of academic ability have been administered to a student and the results on the psychograph are contradictory. The counseler interprets the contradictory findings in terms of his knowledge of the validity, reliability, norms, etc., of the tests in question. For facilitating reference, this discussion of the test results may be divided into the following categories: *General Ability, Achievement, Special Disabilities, Special Aptitudes, Interests,* and *Personality.* The following illustrates one treatment of test results:

> *Test Analysis:*
>
> *General Ability*—Three tests of college aptitude show this boy to be poor material for college. Only on the arithmetic sub-test of the ACE test does his performance exceed the median.
>
> *Achievement*—High School achievement rather high according to report card. Mostly A's and B's. Standardized achievement tests, however, show no percen-

tiles higher than 60, mostly below 40 for freshman norms. High school grades probably a result of very pleasant, outgoing personality.

Special Disabilities—Reading speed and comprehension, average. Spelling very weak.

Special Aptitudes—Clerical test placed him in upper decile on clerical worker norms. Other special tests administered but considered irrelevant.

Interest—Claimed interest, medicine. Strong Vocational Interest Blank shows only one primary pattern—Group VIII—Business Detail occupations. Occupational level is low, only 31 percentile. This contraindicates professional training, especially in view of general ability scores. No verification for claimed interest in medicine.

Personality—Bell Adjustment Inventory reveals no deviate scores. Social scale indicates an aggressive lad. (Clinically verified.)

It often happens that in the process of dictating or writing the test analysis, certain clinical material comes to light which was omitted from the *Clinical Data* paragraph. This may be dictated after the test analysis. Suppose, for example, the personality tests indicate a neurotic individual. The following notation might be appropriate:

Additional Clinical Data: While the interview was in progress, Elmer continually toyed with an ashtray that was on my desk. He frequently twitched his shoulders, bit his upper lip many times. His movements were quick and jerky; he spoke in a nervous, hurried manner. This would tend to verify the scores on the Bell Adjustment Inventory.

Having analyzed the test results and other data in terms of the student's problem or complaint, we are ready for a diagnosis. The counselor's statement of the problem may or may not agree with the student's statement. The student may make, let us say, a diagnosis of under-achievement where the counselor will write:

Diagnosis: Emotional problem. Probably a result of level of aspiration which is much too high for level of

ability. No real disjunction between ability and achievement.

Another example of a counselor's diagnosis follows:

Diagnosis: Under-achievement. Probably due to lack of motivation, lack of study habits, too much social life. Parental involvements. (No identified problems in health area or financial area.)

It is important to know whether the counselor has investigated other possible areas of diagnosis. If he has inquired into the health status of the individual and learned of no positive findings, the case notes should contain a brief statement to this effect. Thus, the absence as well as the presence of problems is noted. This statement (under *Diagnosis*) further illustrates the point:

No problem identified in health area. Physician's report enclosed.

Having stated a diagnosis, what is the next step? Logically, the treatment (or counseling) phase of the total process. This area requires especially careful reporting. The counselor must try to delineate the techniques that he used in treating the student for his problems. He should state what he tried to do, and the methods he used to achieve his purposes. If he tried to modify the individual's attitudes, he will mention what facts were brought up, what emotionally colored words and cliches he used and so on. If he listed a series of facts for the student to consider, these should be summarized in the record. The plan of action accepted by the counselee should be reported, if possible, in 1–2–3 order.

If treatment is directed at more than one problem, the notes should be organized accordingly. As an illustration suppose the diagnosis indicates an emotional problem arising out of a history of social backwardness, and an educational problem of unwise curricular choice. The case notes would be organized as follows:

Counseling: (or Treatment)

Emotional Problem: Part of the treatment was centered around catharsis. Frank seemed to get relief in talking out his problems to an interested and sympathetic listener. He told me of numerous occasions where he had been miserable and unhappy because he never had learned how to meet people (more of this under Clinical Data). My first step was to have him state that this personality trait was undesirable for life adjustment and that steps should be taken to correct it. My question was: "Do you want to remedy this defect in an otherwise pleasant personality?" He answered in the affirmative. I outlined the following steps for rehabilitating him socially: (1) Register in the Social Skills Course of the Y. M. C. A. (2) Attend meetings of the Cosmopolitan Club where he could use his knowledge of world events. (3) Participate in the Y. M. C. A. mixed group activities. (Letters were sent to the appropriate officer of each group to insure a personalized reception.)

Scholastic Problem: My job was to dissaude him from continuing in pre-business and in having him accept a substitute program of general education. First I pointed out the standard of competition in the professional School of Business. This made no dent in his armor. He still maintained that his D+ average would come up to a C this year. Knowing his dislike of courses involving math, I showed him the courses as catalogued in the professional business curriculum: statistics, finance, money and banking, theoretical economics, insurance accounting, and so on. (With silent apologies to my friends who teach these courses) I told the student that these courses were "highly theoretical and abstract" and considered "very dry." On the other hand, courses in general education were practical and interesting; no economics or math prerequisites were necessary. I described some of the interesting features of the Orientation Courses. He finally agreed to think it over. I outlined this plan of action: (1) see the counselor in the general education unit for further information (I arranged an appointment); (2) discuss the matter with his folks; (3) secure transfer blanks from the Registrar's Office.

The next step follows logically, what is the prognosis, the expectation? What results may we reasonably expect if the student accepts our advice? if he acts contrary to our advice? A prognosis should be stated for each major alternative open to the student. For example:

> *Prognosis:* In General College, good. About C + or B average for year. In pre-medicine, failure after one or two quarters. Prognosis reserved on emotional problem.

Frequently a prognosis is stated conditionally. For instance, if a family problem complicates the case, the prediction of adjustment in another area is stated with a proviso:

> *Prognosis:* Educational adjustment satisfactory providing mother's attitude can be changed, or student moved to a campus residence.

Finally, in order to verify the diagnosis or to continue with the treatment, and to learn the effectiveness of our counseling, as well as the accuracy of the prognosis, we arrange for a follow-up. The case notes should state clearly the follow-up plans of the counselor:

> *Follow-up:* Send notice to return for further counseling in two weeks. Determine whether student has attended group activities; inquire about parents' reactions; get grade report.

The same general procedure, with modifications, is used in recording subsequent interviews. Counselors should dictate progress reports as new data come to light in the development of the case.

> *Progress Report:* Three time distributions show study hours of 20, 24 and 31 respectively for the first three weeks of the quarter. Waste time is reduced from 18 to 10 to 6 hours. His comment, "Now there is so little time in which I can do nothing." I encouraged him to continue the good work. He says he enjoys chemistry very much.

Additional Clinical Data: I asked Frank what was driving him to do such good work. He analyzed it after several halting attempts in this way: (1) for the first time he was interested in school; (2) he feared that he would have to return to the pattern-maker's bench if he failed; (3) he felt he owed me something for the trouble I had gone to in getting him re-admitted.

V. CONCLUSION

From the foregoing it may appear that counseling follows a prescribed routine. This is far from true. In the actual counseling situation, a shifting attack generally prevails. Now the counselor is interpreting test data, now he asks a question, now he gives advice, now he makes a diagnosis, and so on. The counseling process is a dynamic, fluid process. Case notes cannot adequately reconstruct the counseling process. No matter how facile the interviewer with the spoken or written word, he can never reproduce the events of the counseling situation so that the reader will perceive the same total situation. Photographic and chronological reproduction being impossible and demonstrably unnecessary, we should recognize that verisimilitude must serve as an appropriate standard for reporting.

To summarize: We have presented six functions of case notes in psychological counseling. We have briefly described the characteristics of desirable case notes which meet our stated objectives. And finally, we have demonstrated the form and content of case notes now in use at the University of Minnesota Testing Bureau.[4] Our method provides for a statement of the problem as perceived by the client or referring agency, meaningful clinical data, interpretation of test results, diagnosis, treatment, prognosis, and follow-up. Psychologists will immediately recognize these as the major steps in the clincal method.

[4] The steps in clinical counseling which precede the recording process are described in Williamson, E. G., and Darley, J. G., *Student Personnel Work*, New York, McGraw-Hill Book Company, 1937.

FIFTH REVISION OF KUHLMANN–ANDERSON TESTS*

ROSE G. ANDERSON
Psychological Service Center, New York

THIS revision was undertaken chiefly for the purpose of increasing the range of mental ages possible in the Grade II booklet and in the Grades VII–VIII booklet. Consistent use of the tests had shown that superior children tested by these booklets had less opportunity to exhaust their abilities than in the case of the other booklets. Since the upper tests in the Grade II booklet are also used in Grades III and IV and tests in the Grades VII–VIII booklet are also used in the grades above and below these, the change of tests in these two booklets affects materially the entire range of tests. It also necessitates testing at all ages in the school population. Additional changes were therefore made in certain tests throughout the scale.

The table below summarizes the changes, showing the num-

	Higher M.A. for	Maximum M.A.'s	
		Old	New
Grade I—First Semester	5 tests	7–6	7–8
Grade I—Second Semester	7 tests	7–8	8–1
Grade II	9 tests	9–6	10–11
Grade III	7 tests	11–5	13–3
Grade IV	5 tests	12–9	15–4
Grade V	4 tests	15–4	18–0
Grade VI	5 tests	17–10	19–11
Grades VII–VIII	4 tests	20–4	25–2
Grade IX–Maturity	4 tests	30–5	40–0

* This paper was read at the meeting of the Eastern Psychological Association, Atlantic City, April 5, 1940.

ber of tests in each booklet for which higher mental ages are now attainable and the new maximum attainable mental age on the booklet compared with the old.

The higher mental ages are due in the majority of cases to added trials in sixteen tests. In other cases they have resulted from the shifting of tests in arriving at the present order. Two new tests have been added (new numbers 24 and 38). Two old tests were dropped (old numbers 17 and 27). Two additional new tests were tried out but not incorporated, a modification of test 27 and an adaptation of the Kuhlmann test of inverted lines. The following tests (old numbers) have added trials: 3, 6, 7, 8, 11, 12, 13, 15, 16, 19, 21, 25, 34, 36, and 38. The items in test 9 have been made more difficult. Added trials are of increased difficulty in the majority of tests. Exceptions to this are tests (old numbers) 13 and 34.

Further indicated changes were made in the content or arrangement of several additional tests (old numbers 1, 2, 5, 8, 29, 32, and 39).

Departures were made from the old arrangement of tests to eliminate from the lower ages or to move to higher ages tests dependent on school training. Tests (old numbers) 12 and 13 which require writing of numbers were moved from Grade I—Second Semester to the Grade II booklet. No other tests below the Grade III booklet are now dependent on specific school training. Test 17 in the Grade II booklet requiring reading of words was dropped. The first test of the latter nature is now the last test in the Grade III booklet. Half of the tests in the Grade III booklet make no demands on reading or writing.

The changes in the Grades I and II booklets are more significant than the above increases in maximum M.A. indicate. The situation was that some children of superior intelligence had not been able to attain the M.A. previously possible on these booklets because of the dependence of tests 12, 13, and 17 on writing numbers and reading words because of limited school experience in these respects.

The increased range of mental ages has been added with no increase in the time required for administration in the upper age booklets. The added trials in these tests are for the more adequate measurement of superior students who have previously been able to cover all items in the tests in question before the expiration of the allotted time. The added time for the lower age booklets is from two to four minutes. The increase is in those tests in which each item is administered individually.

Instructions for giving have also been revised in some instances. The drawings for several tests have been improved.

A simplified method for securing the median mental age (or mental growth units) has been provided.

The preliminary try-out of the changes was made in a suburban New York school with a population of about 1100. Indicated changes were then made and the tests given in three Minnesota towns, considered comparable to those from which the original norms were obtained. These supplementary norms are based on 3074 cases from these Minnesota schools, 200-368 cases for each age group from six through seventeen. As in the case of the original norms, the tests were administered by one of the authors (Anderson), assisted by trained psychologists working under her direct supervision.

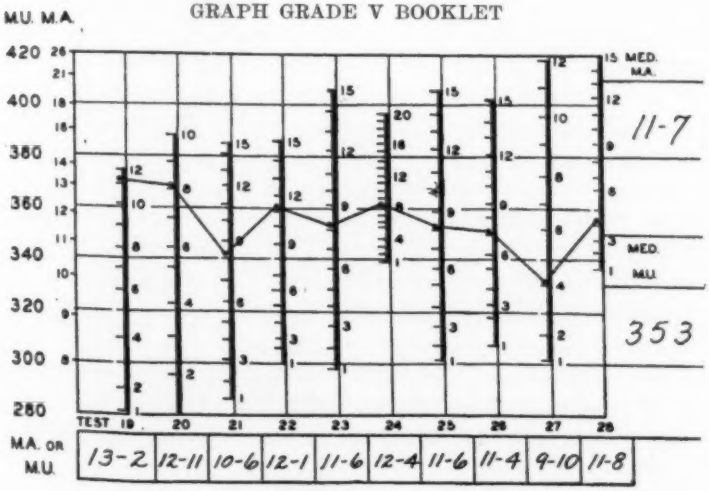
The practice recommended by the authors of fitting the test booklet to be used to the expected level of the group to be tested has proved satisfactory for groups in general. This practice, however, does not insure the accuracy of the ratings of the few pupils definitely inferior or superior to the median of the group. These children, as has been previously recommended, should be selected on the basis of the numbers of their zero and maximum scores. For the greatest accuracy, those receiving two or more zero scores should be given the tests necessary to be rated on the next lower booklet. Correspondingly those receiving two or more maximum scores should be given the supplementary tests necessary to be rated on the next higher booklet. Only thus can all children in a

group be adequately measured. The failure to make this follow-up deprives users from reaping the full benefits from the tests.

However, at one point in the Kuhlmann-Anderson scale this practice has been less feasible. This was in the case of superior first and second graders and for the reasons indicated above. The changes incorporated in the Fifth Revision now make adequate provision for exhausting the ability of the very superior first and second graders, individuals or groups, by using a higher booklet.

A profile graph has been devised which shows at a glance the number of maximum and zero scores. This provides a direct basis for determining whether the most appropriate booklet has been used. Examination of this graph also facilitates the interpretation as to whether failures or successes in specific tests may be related to the nature of the content, *e.g.* reading, non-reading, numbers, etc. While recommended for the above purposes, the profile should not be used to draw conclusions from the results of the individual tests with respect to specific abilities presumed by the user to be tested by the separate tests.

GRAPH GRADE V BOOKLET



The graph for the Grade V test booklet is inserted above, illustrating its use in an individual case. The heavy vertical lines show the range of application of the separate tests. The short horizontal lines on the vertical lines mark off the separate trials. Dots are made on the appropriate horizontal lines corresponding to the number of trials passed in each test. The irregular line connecting the dots shows at a glance that the battery of tests used is appropriate for the child in question, since the performance is in general in the middle range tested by the booklet with no zero or maximum scores.

Further examination of this graph raises the question whether there is a relation between level of performance and dependence upon reading ability. The two highest M.A.'s are earned in tests which do not require reading and the lowest M.A. is earned on the one test in the battery which requires sentence reading. The need to find out more about this pupil's reading ability is definitely indicated.

The statement has been made that the Kuhlmann-Anderson Test in common with other group intelligence tests measures achievement rather than intelligence. The statement has also been made that increasing percents passing a test at successive chronological ages are not conclusive evidence with respect to validity since other attributes, *e.g.* height, also increase with age but are not correlated with brightness.

Supplementary to the discussion in the Kuhlmann-Anderson Instruction Manual relative to these matters we present here some results for selected tests 14, 32, and 39 (new numbers) to see what bearing they may have on the above statements. The number of passes in Test 32 depends upon range, variety, and specificity of information. It might therefore be considered an achievement test rather than an intelligence test.

Table I shows appreciable increases from age to age in percents passing a given number of trials. It also shows consistent decreases in each age in percents passing an increas-

TABLE I

TEST 32

Percents Passing

Trials	Ages	13	14	15	16	17
	Cases	250	368	312	332	287
6 or more		55.2	67.7	79.2	83.7	85.7
8 or more		22.0	36.7	51.0	53.6	61.7
10 or more		6.7	13.0	24.4	26.5	32.8
12 or more		1.7	2.8	4.8	7.8	11.8
14 or more		.0	.3	.3	.0	1.0
15		.0	.0	.0	.0	.0

ing number of trials. Does this mean anything beyond the fact that fifteen year olds have been in school a year longer than fourteen year olds and have therefore had more opportunity to acquire the needed information? This would be the interpretation of some.

TABLE II

TEST 32

Median Trials Passed

Age	Grade VI	Grade VII	Grade VIII	Grade IX	Grade X
11	(119)* 5.73				
12	(92) 4.87	(131) 6.20			
13		(92) 5.53	(111) 7.56		
14			(129) 6.68	(194) 7.82	
15				(141) 7.56	(134) 9.0
16					(121) 8.0

* Figures in parentheses indicate number of cases.

But let us examine Table II. Curiously enough when school grade is held constant, we find that younger children in any given grade get higher scores than those a year older. Their relatively superior performance apparently cannot be attributed to specific school training, since many of the

younger ones have had a shorter school experience. This would be more strikingly shown if we used children of two years difference in chronological age. We have used the one year difference because the numbers are large enough to be significant and because the evidence is more convincing when the comparison is severe.

If, on the other hand, the increases in Table I from age to age are attributable to general maturation factors, why is it that the children who have had a year more opportunity to mature (Table II) do less well? Also, in view of such results, can the higher scores of twelve year olds in the seventh grade over the twelve year olds in the sixth grade be accounted for in any other way than that of the mental superiority of the former? Is it debatable that the same factor which has resulted in rapid school progress also results in a better test showing?

TABLE III

TEST 39

Percents Passing

Trials	Ages	13	14	15	16
	Cases	238*	368	312	332
5 or more		58.9	76.7	86.3	87.4
10 or more		15.1	17.2	31.3	39.5
15 or more		1.2	2.5	4.8	10.8
20		.0	.0	.0	1.5

* Does not include thirteen year olds in Grade V.

Similar results are presented for Test 39. There seems little basis for accounting for performance in this test on the basis of any specific training, yet there is no apparent difference in the trend of the results in Test 32 as compared with Test 39.

Results for Test 14 are presented because performance in the test depends upon rapid counting and writing of numbers up to 14. It is the one of the lower age tests most susceptible to the criticism that intelligence is not being measured.

TABLE IV

TEST 39

Median Trials Passed

Age	Grade VI	Grade VII	Grade VIII	Grade IX	Grade X
11	(119)* 4.29				
12	(92) 3.69	(131) 5.4			
13		(92) 4.84	(111) 6.99		
14			(129) 6.34	(194) 7.78	
15				(141) 7.88**	(134) 9.27
16					(121) 8.25

* Figures in parentheses indicate number of cases.

** Only exception to younger children in same grade getting higher scores.

In the case of younger children, there is undoubtedly a factor of both physiological maturation and of specific experience in doing this test. Yet younger children in the same grade compare favorably with children a year older. The evidence is that something beyond these two factors is being measured by this test.

TABLE V

TEST 14

Percents Passing

Trials	Ages	7	8	9	10
	Cases	217	223	235	220
11 or more		9.2	35.9	66.0	82.3
12 or more		4.6	22.0	50.2	67.3
13 or more		.5	10.3	29.8	50.5
14 or more		.0	3.6	14.9	35.5
15 or more		.0	.9	7.2	19.1
16		.0	.5	3.0	11.8

Essentials in assuring the highest possible validity of tests include (1) critical original selection of tests, (2) avoidance of tests dependent on special training, (3) placing tests dependent on general training where children have had equal

TABLE VI

TEST 14

Median Trials Passed

Age	Grade II	Grade III	Grade IV	Grade V
7	(136)* 9.05			
8	(95) 8.50	(110) 11.39		
9		(95) 11.39	(113) 12.79	
10			(100) 12.85	(101) 13.53

* Figures in parentheses indicate number of cases.

opportunity for such training, (4) a sufficiently large number of likely tests so that only those having the highest discriminative capacity need be retained, (5) retention of those which discriminate best between unselected successive age-groups, (6) further detailed analysis of test results to substantiate the indication that the tests measure something beyond general training, *i.e.*, that they measure the capacity to acquire it.

Adequate use of tests involves always a critical analysis of each child's performance to determine whether and, if so, in what respects the limitations of the individual child's experience or his handicaps render the usual generalizations invalid. That such exceptions exist should always be recognized. That they do, does not invalidate these generalizations for the great majority of children.

A COMPARATIVE STUDY OF THREE TESTS OF VISUAL ACUITY

GEORGE SPACHE

Brooklyn Friends School, Brooklyn, New York

SINCE its standardization in 1862 by Professor H. Snellen, the Snellen Chart has become the most commonly used instrument for the testing of visual acuity. Probably because of its ease of administration, it has been the standard vision test used in American schools. It is the purpose of this article to point out some of the limitations of this test, and to compare the results of its use with those obtained from several more recent instruments.

The following criticisms have been made of the Snellen Chart:

1. It is questionable whether the reading of the letters and nothing else is a sufficient test of visual acuity (7, 5).
2. Only 20–40 per cent of the children are identified who need the help of an eye-specialist (2) and a large percentage of the children who have 20/20 vision according to the Snellen Chart are found to have other defects (9).
3. The test does not appraise the efficiency of the eyes at reading distance (10 to 16 inches) (1).
4. The test does not appraise the coordination of the two eyes (2, 3). For since each eye is tested separately, there is no measure of the efficiency of two-eyed vision or of the coordination of the eyes in vision. It is a common occurrence for children to suppress the vision in one eye and to use the other to excess.
5. The test is not critical in detecting astigmatism because an individual can with conscious effort force his eyes to read small enough type to pass the test even though his refractive error is such as to make sustained reading difficult and uncomfortable (2, 3).
6. The test provides no measure of the degree of far-sight-

edness, eye-muscle imbalance, limitations of the visual field or aniseikonia.

7. The test type is not sufficiently varied, so that it is possible for subjects to memorize some of the lines (7).

8. Some of the lines are easier to read than others, while some are especially difficult by reason of the similarity of the letters (7).

With the assistance of Mrs. Dulcie Veit, the school nurse, and the cooperation of Dr. C. H. Laws, the school physician, the writer was able to use the Snellen Chart, the visual acuity test of the Eames kit (4) and the Jensen Tests for Color-Blindness, Visual Acuity and Astigmatism (8) with a number of elementary and secondary school children. The following data were derived from this testing:

TABLE 1

Comparison of the Snellen, Eames and Jensen Vision Tests

Number	Grades	Tests	Correlation*
81	I to XII	Snellen-Eames Visual Acuity	.506
80	I to XII	Snellen-Jensen Visual Acuity	.693
113	K to XII	Eames-Jensen Visual Acuity	.609
111	K to XII	Eames-Jensen Astigmatism	.334

* These correlations are C or coefficients of contingency. In the 4×4 grouping used here, C cannot exceed .866 and is only roughly equivalent to the Pearson product-moment r. (6, pp. 195 ff.)

The card for testing visual acuity of the Eames kit is quite similar to the well-known Snellen Chart. It consists of a line of 6 capital letters and a line of capital E's. The test is administered with the subject at a distance of 20 feet. Each eye is tested separately by covering the other with a card or paper shield. In addition, the examiner's kit includes cards for testing coordination or fusion and astigmatism and a lens for determining the presence of far-sightedness. The Jensen test consists of four color plates, similar to the Ishihara Color Perception test for determining color-blindness, and six circular grids or circles composed of black bars. In testing

visual acuity, the subject is placed at a distance of ten feet. From this point, the subject is moved backward or forward until he is able to see the direction of at least one of the grids. The same procedure is followed in testing both eyes and each eye separately. Defective acuity may be expressed in fractional parts of normal vision or in Snellen units.

TABLE 2

Correspondence of the Snellen, Eames and Jensen Vision Tests

Visual Acuity SNELLEN	EAMES			
	Both normal	Right defective	Left defective	Both defective
Both normal	71	1	0	2
Right defective	1	0	0	0
Left defective	0	0	0	0
Both defective	2	1	0	3

Visual Acuity SNELLEN	JENSEN			
	Both normal	Right defective	Left defective	Both defective
Both normal	72	0	0	0
Right defective	1	1	0	0
Left defective	0	0	0	0
Both defective	4	0	0	2

Visual Acuity EAMES	JENSEN			
	Both normal	Right defective	Left defective	Both defective
Both normal	99	1	0	0
Right defective	3	0	0	0
Left defective	2	1	0	0
Both defective	4	0	1	2

Astigmatism EAMES	JENSEN			
	Both normal	Right defective	Left defective	Both defective
Both normal	42	0	1	2
Right defective	8	0	0	0
Left defective	5	1	0	0
Both defective	41	2	0	9

Table 2 is read as follows, "In the comparison of the Eames and Snellen tests of visual acuity, 71 cases were found normal in both eyes according to both tests, one child was found defective in the right eye by the Eames test but declared normal in both eyes by the Snellen Chart. Two cases found de-

fective in both eyes according to the Eames were normal in both eyes according to the Snellen."

In the first three charts, the tests are compared in their testing of visual acuity or keenness of vision. The lower right-hand chart is the comparison of the Eames and Jensen results in the testing of astigmatism.

In computing the correlation between the results of a pair of tests, a 4×4 grouping was employed. The classifications were (1 both eyes normal, (2 right eye defective, (3 left eye defective and (4 both eyes defective. Thus it is possible to determine the extent to which the results of a pair of tests coincide in any of these four groups. In the correlation of .506 between the Snellen Chart and the Eames test, 71 children were found to have normal vision in both eyes and 3 were defective in both eyes according to both tests. In other words, the results of the tests were identical in 91 per cent of the 81 cases. Results were directly contradictory in that one test found the children defective in one or both eyes and the other found them normal in only 6 cases. The similarity of the nature and administration of these tests would lead the examiner to expect even greater similarity in results than these findings indicate.

In the correlation of .693 between the Snellen Chart and the Jensen test for visual acuity, identical results were obtained for 75 or 93 per cent of the cases. Contradictory results occurred in only 5 cases. Despite the distinctly different nature of these tests, they are decidedly more similar in results than the Snellen-Eames pair. In the comparison of the Eames and Jensen tests, results were identical in 101 or 88 per cent of the cases. Results were contradictory in 11 cases. This correspondence may be expressed in a C of .609.

The validity of the three tests, or the extent to which they really measure visual acuity, may be judged by the extent to which their results are similar. The average C of each test with the other two is Jensen – .651, Snellen – .599 and Eames – .557. These validity coefficients lead to the conclusion that

the Jensen visual acuity test is superior to either the Snellen or the Eames. The superiority of the Snellen over the Eames may well be due to the greater amount of testing in the former. There are 7–8 lines of type for testing in the Snellen and only a single line in the Eames.

The Jensen test is also superior to the Snellen in that it is not subject to a number of the limitations of the latter. Specifically, it provides a measure of two-eyed vision and of astigmatism, as well as visual acuity. It is not composed of letters which may be memorized or which differ in ease of visibility. It is easy to administer, even with five year olds, and moderately priced. These points, added to its superior validity, warrant its use in place of the Snellen Chart.

Since both the Eames and Jensen tests permit the testing of astigmatism, a further comparison is possible. The Eames card for testing astigmatism is a half-wheel composed of heavy black lines. The subject is placed at a distance of 20 feet and asked, "Are all these lines the same?" or "Are some of these lines darker than the others?" Astigmatism is tested by the Jensen test by determining the difference between the point at which the subject can just see one of the grids and the point at which he can just see all of them. The direction of the grids which are most difficult for the subject to see indicate the axes of his astigmatism.

From the testing of 111 children, a C of only .334 was obtained from the two tests for astigmatism. Identical test results were found in only 51 cases. Forty-one cases were found normal by the Jensen but defective in both eyes by the Eames. Only two children were defective in both eyes according to the Jensen and normal according to the Eames. The remaining 15 gave results only slightly less contradictory. These results can have but two possible meanings. Either the Jensen test is failing to detect astigmatism or the Eames is indicating astigmatism where there is none.

The nature of the questions which must be asked in the administration of the Eames test is conducive to either positive

or negative suggestibility, particularly with the younger children. It was apparent that the children were attempting to answer the questions in a manner pleasing to the examiner or in the manner which they thought was correct. This difficulty is not present in the use of Jensen since the task is merely one of indicating the directions of the grids. This element of suggestibility has apparently invalidated the Eames test as a real measure of astigmatism and rendered it unsuitable for school use.

BIBLIOGRAPHY

1. BETTS, EMMETT ALBERT, "Reading Disability Correlates," *Education*, 56, September 1935, pp. 1–7.
2. BETTS, E. A., *The Prevention and Correction of Reading Difficulties.* Evanston, Illinois: Row, Peterson, 1936. Pp. 402.
3. EAMES, THOMAS HARRISON, "Improvement in School Eye Testing," *Education*, 56, September 1935, 14–18.
4. EAMES, T. H., *Eames Eye Test.* Author: 31 Chester Street, West Somerville, Mass. Price $3.00.
5. FARRIS, L. P., "Visual Defects as Factors Influencing Achievement in Reading," *Clearing House*, 9, December 1934, 226–228.
6. GARRETT, HENRY E., *Statistics in Psychology and Education.* New York: Longmans, Green, 1926. Pp. 317.
7. GARRETT, HENRY E. and SCHNECK, M. R., *Psychological Tests, Methods and Results.* New York: Harper and Bros., 1933. Pp. 235.
8. JENSEN, MILTON B., *Tests for Color-Blindness, Visual Acuity, Astigmatism.* New York: Psychological Corporation, 1935. Price $3.00.
9. KEMPF, GROVER A., JARMAN, BERNARD L. and COLLINS, SELWYN D., "A Special Study of the Vision of School Children," *Public Health Reports*, 43, pt. 2, no. 27, 1713–1739. Washington, D. C.: Government Printing Office.

AN EXPERIMENTAL STUDY OF FOUR METHODS OF MEASURING UNIT COSTS OF OBTAINING ATTITUDE TOWARD THE RETAIL STORE[1]

JOHN W. HANCOCK
Purdue University

STATEMENT OF THE PROBLEM

THE general problem under investigation may be stated as an attempt to determine methods which will most efficiently measure the attitude of large populations toward the retail store. The retail store may be conveniently divided into two general parts: the independent store and the chain store. In this paper the author does not intend to intimate what the attitude should be toward either chain or independent store. The results shown herein are experimentally determined and tend to point out attitudinal relationships together with certain methods for obtaining the specified attitude from large geographical areas.

THE MEASURING INSTRUMENT

After a careful survey of the literature in the field of attitudes and attitude measurement[2] it was concluded that a modified form of the five-point rating scale would best serve the purposes of this study.

[1] This work was done under the direction of Dr. H. H. Remmers, to whom my appreciation is hereby expressed. A more complete report of this study is on file as a doctoral dissertation in the Purdue University Library. This thesis is partial fulfillment of the requirements for the Ph.D. degree at Purdue University.

[2] A bibliography of 147 titles is included in the thesis on file in the Purdue University Library.

In order to understand the attitudes of the people toward chain and independent stores, the variables of the situation must be investigated.

From a preliminary survey of field literature it was decided that public attitude toward the retail store could be measured by an appropriate instrument which would investigate the following areas:

1. Merchandising
   a. Quality of goods
   b. Style of goods
   c. Completeness of stock
2. Material services rendered
   a. Money savings
   b. Correct change
   c. Sufficient number of clerks
   d. Sufficient floor space for customers
   e. Parking space
   f. Delivery service
   g. Charge accounts
3. Personnel of stores
   a. Dependability
   b. Courtesy
   c. Salesmanship
   d. Friendliness
4. Advertising
   a. Integrity of advertising
   b. Window displays
   c. Amount of advertising
5. General policies of stores
   a. Integrity
   b. Cleanliness
   c. Refunding of money
   d. Exchange of merchandise
   e. Location of store
6. Relationships to community
   a. Taxation of stores
   b. Benefit of stores to community
   c. Cooperation
   d. Unfair competition

In order to construct an instrument that would quantify attitude reactions, a survey was made of the field of attitudes and attitude measurement.

PRELIMINARY WORK

Methods for Obtaining Response

In order best to determine methods for obtaining responses to requests for individuals to fill out an instrument which would reflect attitude toward the retail store, it was decided that the merits of four methods should be investigated: (1) A method of sending out by mail the instrument with an appropriate letter explaining the purpose of the study and instructions for marking the instrument; (2) A method of sending out by mail the instrument with the inclusion of twenty-five cents to be kept by the addressee for his efforts used in filling out the instrument. An appropriate letter was also included which explained the purpose of the study, instructions for filling out the instrument, and an explanation for the inclusion of the twenty-five cents; (3) A method of sending out by mail the instrument with an appropriate letter stating the purpose of the study, instructions for filling out the instrument, and a statement that upon receipt of the filled-out instrument twenty-five cents would be forwarded to the addressee; (4) A method of personal interview in which the two investigators of the study contacted personally the desired number of individuals. The same instrument was used in the personal interviews as was used in the three preceding methods. These methods are hereafter referred to as Method 1, Method 2, Method 3, and Method 4 respectively.

Upon establishing a list of names of individuals to contact in this study, the list of names being obtained from the personal tax records of the county assessors, four groups were chosen at random from the list for each county. Each group was to be approached by only one of the above methods. The method of establishing these lists will be explained in detail later.

Just as it is necessary to make a diplomatic approach to an interviewee, it is important to approach the addressee by mail with a letter that is not only straightforward and sincere, but forceful and moving. For this reason much attention was given to the composition of the explanatory letters accompanying the instruments.

Since there were three mailing methods it was necessary to construct three different letters that explained the purpose of the method by which a particular individual was approached.

Great efforts were made to word the letters so that everyone would be able to understand the contents, so that the letters would not be of unnecessary and tiring length, and so that they would be tactful and forceful in urging the individual to cooperate in the study under investigation.

Enough has been said about the first three of the methods. The fourth method warrants some additional attention. Since variables of personality of the interviewer go into the conducting of an interview, and since two interviewers were used, it was thought best that the interview technique be made as standard as possible. Also, since it was desired to compare results from the interviews with results gained through the other three methods, it was thought that the personality of the interviewer should be kept as much in the background as possible. Thus each interviewer in conducting his interviews would ask the interviewee the questions just as the questions appeared on the instrument. Then the interviewer would mark the instrument according to the response of the interviewee. For easy reference, the two interviewers will be called "G" and "H" in this study.

Selection of the Experimental Territory

It was desired to establish a territory that would be fairly representative of the state of Indiana with the exception of the district about Lake Michigan. An arbitrary line of reference was drawn from Lafayette, Indiana to Cincinnati, Ohio. It was necessary to set up an area that would have

rural and urban population comparable to the state-wide population ratio, as well as an area to which the various methods of approach could be applied within time and expense limits. It was found from the late census data that a territory roughly twenty miles on each side of the reference line would satisfy the requirements.

With this general area in mind the attention of the workers was turned toward the features of the several counties within the area. The counties are Boone, Clinton, Fayette, Franklin, Hamilton, Hancock, Henry, Madison, Marion, Rush, Tippecanoe, Tipton, and Union.

Figures were obtained for the population of each county, its percentage to the total population for all thirteen counties, rural and urban percentages of population of each county, and the proportional number of returns that it would be necessary to obtain from each county in order to total 1500 returns, which number was thought to be sufficient for subsequent analyses.

In order to justify estimates made, a preliminary experiment was set up in Tippecanoe County. It was estimated from results of previous allied studies that returns should approximate the following: Method 1, 20 per cent; Method 2, 60 per cent; Method 3, 60 per cent; and Method 4, 80 per cent. The percentage returns from Tippecanoe County proved to be as follows for the methods respectively: 21, 52, 44, and 74 per cent.

The per cent of rural and urban population in the various counties, the number of names to be drawn for each county to yield the desired number of returns, the number of rural and the number of urban names drawn, and finally the number of rural and urban returns desired from each county to contribute proportionally to the desired total of 1500 returns were determined from these percentages.

Safety factors in terms of additional names were added to prevent the necessity of returning to various county seats, should our percentages prove at fault.

A second preliminary experiment was carried out to determine the advisability of using permit return envelopes. This experiment was set up for Madison and Henry counties. Letters sent to 119 rural and 491 urban individuals showed the cost per unit return by the permit return method to be 18.3 cents as against 33.7 cents for the stamped envelope return. It was concluded that the use of the permit return method should be employed.

COST ESTIMATES AND FINAL COSTS

Cost Estimates

The preliminary investigation in Tippecanoe County showed that the unit costs of usable returns were, for the four methods: Method 1: 43 cents; Method 2: 57 cents; Method 3: 50 cents; and Method 4: 19 cents.

On the basis of information in the hands of the workers at this point, a cost estimate was made for the entire study. The estimations were as follows:

Cost of obtaining sampling group	$ 200.00
Cost of obtaining 375 completed interviews based on an estimated 75 cents per interview	271.25
Cost of 375 units for Method 1, based on use of permit mailing costs at 24 cents per unit as found in the preliminary study in Madison and Henry counties	90.00
Cost of 375 units for Method 2, based on 14 cents per unit return plus 25 cents in each of 862 units	361.75
Cost of 375 units for Method 3, based on 42 cents per unit return	157.50
Total	$1080.50

To the above total cost would have to be added the cost of clerical labor.

Final Costs

After the study had been completed in its present form, the total costs were found.

Since it proved more convenient and less expensive to obtain a mailing list and complete the interviews within a single county before moving on to the next county, the costs in the following schedule for items marked "Travel", which include all transportation and living costs for the two workers, and that marked "Rental of Cars" should be proportioned among the methods. The schedule of costs is as follows:

Travel	$ 288.33
25 cent pieces	253.00
Rental of cars	148.25
Clerical help	286.47
Stamps and supplies	339.31
Total	$1315.36

As will be noticed, this total exceeds the preceding estimate by the approximate amount of the cost of clerical labor.

The costs were then apportioned to the several methods, the interview method receiving the greatest burden from the travel and car rental expense items.

Method 1	$ 447.30
Method 2	367.22
Method 3	216.21
Method 4	284.63
Total	$1315.36

The above costs are the costs for each method. Table 1, Column 1, shows the number of individuals to whom letters were sent or who were contacted by the various methods. Column 2 shows the number of usable returns obtained by each method.

Column 3 shows in per cents the usable returns and the number of individuals contacted.

The cost per unit of usable returns for each method may be figured from the total cost of each method divided by the

TABLE 1

The Numbers and Per Cents of Usable Returns by Method for Total, Rural, and Urban Populations

	No. Instruments Used by Mail and Interview	No. Usable Returns	Per cent of Usable Returns
Totals Method 1	3726	366	9.56
Rural	540	50	9.25
Urban	3186	316	9.61
Totals Method 2	960	463	47.2
Rural	144	62	43.1
Urban	816	401	47.9
Totals Method 3	1123	204	17.6
Rural	157	28	17.8
Urban	966	174	17.6
Totals Method 4	454	388	85.5
Rural	66	56	84.8
Urban	388	332	85.7
Totals All Methods	6197	1421	22.9
Rural	897	196	21.8
Urban	5300	1215	23.0
Grand Totals	6197	1421	22.9

number of usable returns for each method. Thus the unit cost for each method is as follows:

Method 1	$1.22
Method 2	.79
Method 3	1.06
Method 4	.74

These figures do not include any salary for the interviewers nor for the director of the study, as this would probably vary from one study to another, but include only such costs as seem likely to remain constant for any area of similar nature.

In order to interpret the present costs of the methods used in terms of what it would cost if the study were repeated, the technique of the standard error of a percentage is used. Thus, for Method 1, where the percentage returns was 9.56, the standard error is .458 per cent, which means that the limiting

percentages of returns lie between 8.19 and 10.93. Method 2 shows a percentage return of 47.2. The standard error of this per cent is 1.125 per cent. The limiting percentage returns are 43.82 and 50.58. Method 3 shows a return of 17.6 per cent. The standard error is 1.01 per cent. The limiting percentage returns are 14.57 and 20.63. Method 4 shows a return of 85.5 per cent. The standard error is .66 per cent. The limiting percentage returns are 83.52 and 87.48.

Transforming these percentage limits into limits for unit cost we find that Method 1 would vary between $1.20 and $1.24 per unit. The cost of Method 2 would vary between 76 cents and 82 cents. The cost of Method 3 would vary between $1.03 and $1.09. The cost of Method 4 would vary between 72 cents and 76 cents.

RESULTS

Returns

Table 1 shows a 9.56 percentage return for Method 1, 47.2 percentage return for Method 2, 17.6 percentage return for Method 3, and 85.5 percentage return for Method 4. In each case there is a very slight difference in percentage returns in favor of the urban areas.

Out of the 5,743 letters sent out, 257 letters or 4.5 per cent were returned unclaimed. Of this number, 27 or 10.5 per cent were addressed to rural individuals and 230 or 89.5 per cent to urban individuals. This result is to be expected when the mailing list is obtained as in this study; however, if this could be eliminated, unit costs would be made slightly less.

It is interesting to note in Method 2, that of 960 persons who were sent quarters, 84 or 8.75 per cent of the persons did not cooperate by filling out the attitude blank, but returned the twenty-five cent pieces. Six additional persons returned the quarter as well as the filled-out blanks.

Method Differentials

It is quite important to see whether the methods used to obtain the returns in this study are measuring similar popula-

tions or are obtaining attitude differences which cannot be explained on the basis of chance error. For this reason the results on the four sections were compared by each method for significance of difference with each of the other methods.

In the following discussion it will be well to keep in mind that an attitude measure of 4.00 expresses indifference of attitude. Any measure numerically higher than 4.00 shows favorableness of attitude. Any measure numerically lower than 4.00 shows lack of favorableness of attitude toward the given attitude object. The mean score of Section I, dealing with attitude toward 22 particular chain stores, for each individual was found by adding all of his attitude ratings and dividing by the number of ratings. Distributions of these mean scores gave the mean scores for each method.

By the use of the formula for finding the standard error of the difference between two per cents, the differences of averages among the four methods as applied to Section I of the instrument are investigated. No significant differences are found among the inter-relationships of the various means, the greatest significance ratio yielding 59 chances per hundred that the difference can not be accounted for by chance.

Section II of the instrument attempts to measure attitude toward certain store-tax policies. Methods 1 and 2 show that they measure these attitudes with equal accuracy. Significant differences, however, are brought to light when other intercomparisons of the methods are made. Method 3 measures quite comparably with Methods 1 and 2 on this section. Only two significant differences are shown between Method 1 and Method 3, and only four when Methods 2 and 3 are compared.

When Method 4 is compared with the other methods for this section, eighteen significant differences are found out of 54 possible comparisons. This seems to indicate that the person to person contact has in some way disturbed the interviewee, making him state that the existing taxes are "all right" greatly more often than when the interview method

was not used. This is perhaps the reason that nine of the significant differences noted are accompanied by an average of 11.2 per cent more persons saying "all right" by Method 4 than by any other method.

Of the remaining nine significant differences found when Method 4 is compared with the other methods, eight of them occur when the populations range between 24 and 74 individuals. These populations are too small in this case to bear strict statistical analysis.

Four of the eighteen differences between interviewers are found to be significant. These may be explained on the basis of the preceding two reasons plus the effects of interviewer differences.

Thus, it may be said that Methods 1, 2 and 3 measure Section II with comparable accuracy. Whereas Method 4 measures the contents of this section with markedly different results. The differences are large enough in many cases to be caused by factors other than chance. It has been suggested that the causes may be due to an inadequate population in several cases, but in general due to the face to face contact of interviewer and interviewee.

No significant differences were found among the four methods for Section III of the instrument, dealing with store policies and conditions.

Since no significant differences were found among any of the methods for this section of the instrument, we can say that the differences present can be accounted for by chance and that, as in Section I, no one method samples a population significantly differently from populations sampled by the other three methods. Items 61 and 62 of Section IV of the instrument inquire respectively what per cent of retail stores in the United States should be chain stores and what per cent of retail business should be done by chain stores. No significant differences obtain among the various methods. Thus we are able to state that the four methods sample similar populations and that the differences are chance differences.

Item 63 of Section IV inquires "Do you prefer chain stores to independent stores?" A number of individuals expressed their attitudes by encircling both alternatives and stated that they were indifferent to the question, having no preference as between chain and independently-owned stores. The data showed on the average that 37.25 per cent of the population prefer chain stores to independent stores, 54.55 per cent hold the reverse attitude, while 8.20 per cent state their indifference. Since no difference is significant it is concluded that the methods do not sample out groups that differ significantly one from another in this attitude.

Age Groups Sampled by the Four Methods

The mean ages of the groups sampled by the four methods are not greatly different. The greatest difference is seen to be between Method 3 and Method 4, 7.3 years. No mean difference between the methods is significant. The conclusion may be drawn that none of the methods singles out age groups that are more than chancely different from those represented by the other methods.

From additional data it is found that there are no significant differences between age groups when the mean percentages of income said to be spent in chain stores are compared. This is true for age groupings by five-year intervals between the ages 20 to 70. It is further noted that no significant difference was found when the attitude of those over 40 years was compared with that of all individuals under 40 years of age.

The Per Cent of Spent Income That Is Spent in Chain Stores

In this comparison the rural population reports a 14.98 per cent greater spending in chain stores than does the urban population. This yields a critical ratio of $\frac{14.98}{2.05}$ or 7.25, which means that there are 100 chances in 100 that rural people say they spend a higher per cent of their spent income in chain stores than do urban people.

When the occupational groups are compared the following schedule may be set up with regard to the per cent of spent income said to be spent in chain stores. From low to high the order is best expressed:

1. Owner of Business
2. Salesman, Skilled Worker and Other Occupation
3. Professional Worker, Farmer and Day Laborer

Next let us compare the per cent of spent income said to be spent in chain stores by persons who have gone to grade school but no farther, to high school but no farther, and to college.

The population having attended grade school say they spend a statistically significantly greater per cent of their spent income in chain stores than the population that attended high school. The differences between the grade school and college populations and the high school and college populations are not statistically significant. Thus this factor is apparently not a straight-line function of the degree of education.

Even though the premise is not too well established, the data seem to indicate not only that those of lowest educational status say they spend a significantly higher per cent of their spent income in chain stores than do those of high school education but that this difference is caused by certain factors other than merely educational status. It is perhaps that those of high school education are more influenced by anti-chain store propaganda together with any differences in economies that exist between the two groups.

The "college" group shows only chance differences from the other two groups. This group, with the lowest per cent of spent income said to be spent in chain stores, may also be affected by anti-chain store propaganda; but it is perhaps more logical to assume that the economies of this group permit a greater latitude in its buying power thus making it less vital to make purchases where prices are perhaps lower than some other places.

Item 63 of the instrument asks one to state whether one likes chain stores better than independent stores and to state reasons for the answer. One hundred and thirty individuals from Tippecanoe County answer this question. Thirty reasons supporting chain stores were listed by these 130 individuals with a total frequency of 223. Forty-four statements were made against chain stores with a total frequency of 214.

Attitude toward Chain Stores and Independent Stores

What, finally, does this study point out concerning the attitude of the average individual in the experimental territory toward the chain store and toward the independent store?

Before discussing this attitude let us examine the answers to several items in the scale. The following percentages are based on approximately 1200 cases: 9.24 per cent indicate ownership of a retail business; 19.69 per cent indicate that they obtain a part or all of their income from an independent store; 3.67 per cent state they obtain all or a part of their income from the chain store; 16.59 per cent state that they have worked in a chain store while 39.52 per cent indicate that they have worked in an independent store.

The average attitude toward the twenty-two particular chain stores mentioned in Section I is above the indifference point (4.00) for all methods.

Data concerning the tax issue show: (1) An average of 67.2 per cent of the population thinks the store-tax of $3.50 is "all right" for the independent store; (2) 49.3 per cent thinks the store-tax of the 1st chain store ($3.00) is "all right", with 43.9 per cent thinking this tax "too low"; (3) 55.7 per cent thinks the $10.00 store-tax on the 2nd to 5th chain stores is "all right" while 35.1 per cent thinks this tax "too low"; (4) 53.1 per cent thinks the $20.00 store-tax on the 6th to 10th chain stores is "all right" while 28.9 per cent thinks this tax "too low"; (5) 48.3 per cent thinks the $30.00 store-tax on the 11th to 20th chain stores is "all right"

with a greater percentage thinking this "too low" than "too high"; (6) 39.8 per cent thinks the $150.00 store-tax on all over 20 chain stores is "all right" while 32.1 per cent thinks it "too high" and 28.1 per cent thinks it "too low". In general these data show that about 50 per cent of the population believes that the present store taxes are "all right". With the exception of the $150.00 chain store tax, a higher per cent of the population thinks the chain store taxes "too low" than "too high". This order reverses itself on the $150.00 tax item.

The average attitude toward store policies of merchandising, cleanliness, etc., of Section III, is higher toward these policies in chain stores than toward the same policies in independent stores.

On the average the population believes that 36.69 per cent of the retail stores in the United States should be chain stores. This is followed by data showing on the average the population thinks that 41.5 per cent of the retail business done in the United States should be done by chain stores.

On the average 37.25 per cent of the population prefers chain stores to independent stores, 54.55 per cent prefers independent stores to chain stores, while 8.20 per cent states that it has no preference.

A study of the mean attitudes toward the twenty-two chain stores examined shows that toward these stores there exists a rank order of preference and in most cases the differences between pairs of stores are significant. Sex differences in attitude exist toward only four of these twenty-two chain stores.

Attitudes of Owners of a Business and Others

From a comparison of attitude of owners of retail business and non-owners the following data is drawn.

Data for Section I dealing with attitude toward specific chain stores yield a difference in attitude between the two groups that will exist 100 times out of 100. This difference is in the direction of greater favorableness of attitude toward

the studied chain stores on the part of non-owners of retail business.

Of the eighteen possible comparisons of attitude of store owners and non-store owners toward the tax question (Section II), no difference was found to be significant. The following facts may be noted however: (1) Owners tend more often than non-owners to believe independent stores have too high a store tax; (2) that chain store taxes are too low; (3) between 31.6 per cent and 53.1 per cent of the owners believe the chain store taxes mentioned are "all right", whereas between 41.1 per cent and 73.9 per cent of non-owners believe these same taxes are "all right".

The data for Section III, dealing with store policies, merchandise, etc., yield a significant difference that shows non-business owners more favorable to these attitude objects concerning chain stores than are store owners.

Likewise, data for Section IV show significant differences in the same direction toward what per cent of stores should be chain stores, and what per cent of the retail business should be done by chain stores in the United States.

The percentage of response by store owners for the various methods was investigated. By Method 1, 48 or 12.5 per cent of the responses was from store owners; by Method 2, 56 or 12.0 per cent was from store owners; by Method 3, 22 or 10.5 per cent was from store owners; and by Method 4, 23 or 6.8 per cent was from store owners. Intercomparisons show no significant differences among methods, the greatest significance ratio yielding 80 or 79 chances per 100 that the differences can not be accounted for by chance.

SUMMARY AND CONCLUSIONS

Analysis of the data shows that the four methods in general measure with equal accuracy the attitudes toward the attitude objects used in Sections I, III, and IV of the instrument. These similarities are pointed out through the use of the standard error of the difference technique.

Eighteen significant differences are found from fifty-four

comparisons when Method 4 is compared with the other three methods for Section II. This fact seems to indicate that the personal interview leads the interviewee to state that existing taxes are "all right" greatly more often than when the other methods are used.

It is suggested on the following grounds that Method 2 is the best method to use in an investigation of this kind.

1. It measures with accuracy equalling that of the other three methods the attitude objects in Sections I, III and IV of the instrument.

2. It measures the objects of Section II with the same accuracy as does Method 1 and varies but slightly from that of Method 3.

3. Method 2 costs less per unit return than does either of the other two mailing methods.

4. Method 4 is slightly less expensive than is Method 2; but Method 4 shows such large differences from the other three methods when measuring attitude toward the tax question that Method 2 is judged more accurate than Method 4.

5. With regard to measuring attitude toward any political action taken in the future, the situation would be least similar to Method 4.

6. Method 2 shows higher percentage returns than does any of the other mailing methods and is exceeded only by the interview method.

Analysis of the data has further yielded the following pertinent facts:

1. The average opinion of the 1,299 persons responding is that 36.72 per cent of the retail stores in the United States should be chain stores.

2. The average opinion of the 1,292 persons responding is that 41.50 per cent of the retail store business should be done by chain stores.

3. Of the 457 persons responding, on the average 37.2 per cent prefer chain stores to other retail stores, 7.95 per cent are indifferent, and 54.85 per cent prefer independent retail stores.

4. The three statements that were most often reported in support of chain stores out of 223 reports from Tippecanoe County are: (1) Cheaper in price, 65 times; (2) Better or more complete stock of merchandise, 32 times; (3) Efficiency and promptness of service, 20 times. The three statements most often reported in disfavor of chain stores out of 214 reports from Tippecanoe County are: (1) Certain detriments to the community, 61 times; (2) Operate against the benefit of employee, 56 times; (3) Poor clerking service, 23 times.

5. Certain significant differences exist among the amounts allegedly spent in chain stores by the following seven occupational groups studied: farmer, professional worker, day laborer, salesman, skilled worker, owner of business, and non-classified. The average for all these groups is about 25.6 per cent of their spent income said to be spent in chain stores. The greatest difference exists between the owner of business and day laborer, the latter spending significantly more than the former.

6. The four methods show that the mean age of the population sampled by each is not significantly different and is about 39.65 years.

7. No significant differences exist in per cent of spent income said to be spent in chain stores in persons from 15 to 84 years of age when classed in five year intervals. The per cent is roughly 28.7. Further, no significant difference is shown between age groups below forty years of age and above forty.

8. Of the 1,421 persons responding, 10.39 per cent were farmers, 12.95 per cent professional workers, 12.02 per cent were day laborers, 11.12 per cent salesmen, 20.97 per cent skilled workers, 9.24 per cent owners of a retail business, and 21.99 per cent were unclassified as to vocation.

In conclusion the author wishes to state that no generalizations outside of the data should be made without reservations and that no definite conclusions should be reached from this data for any other geographical territory than the one used in this study.

A COMPARATIVE STUDY OF A SMALL GROUP OF CLINICAL TESTS

HENRY WERNER
Judge Baker Guidance Center

IN a child guidance center, the psychological examiner employs numerous supplementary tests in addition to tests of general intelligence, including tests of school achievement, motor control and manual dexterity, vocational interest, visual and language apperceptions, general information, and aptitude. The supplementary tests serve the purpose of verifying the mental rating and of discovering special abilities or disabilities.

The relationship between the various tests used is not always known. We have selected for study a few of the tests most commonly used in one child guidance clinic, with the purpose of showing the relationship between these tests and comparing our findings with those of previous investigators.

PROCEDURE

The specific group of tests studied include: Stanford Revision of the Binet-Simon Tests, Healy Pictorial Completion II, Trabue-Kelly-Spivak Language Completion, Ferguson Form Boards, and Lowe-Shimberg Scaled Information (1).

Test scores were taken from 392 subjects, including 200 boys, and 192 girls, all of whom were given individually, at the Judge Baker Guidance Center, the complete group of tests named above. In most instances the entire group of tests was given in one interview; occasionally in two interviews not more than a week apart. The subjects were selected to the extent that all defectives, or borderline defectives, psychotic and doubtful cases have been excluded. The group therefore had no defects

which would make test results unrepresentative of true ability.

The chronological ages of the group range from 10 through 17 years. Twenty-five cases were taken from each year in this range, with the exception of 10-year-old girls, at which age only 17 records of the complete group of tests were available. The distribution of intelligence quotients ranged from 80 through 149, the median IQ for boys being 97.8, for girls, 100.1. Approximately 25 per cent of the total cases had IQ's between 80–89; 50 per cent, between 90–110; and the remaining 25 per cent, between 111–149. The distribution of mental ages is shown in Table 1.

TABLE 1A (BOYS)
Chronological Age

M.A.	10	11	12	13	14	15	16	17	Total
8	3								3
9	4	7	1						12
10	4	8	5		1				18
11	4	2	4	6	2				18
12	4	6	2	3	4	4		1	24
13	2	1	2	1	5	5	5	4	25
14	4	1	5	3	5	3	4	4	29
15			4	8	4	5	5	2	28
16			2	3	2	5	5	7	24
17						3	1	4	8
18				1	2		4	3	10
19							1		1
									200
	25	25	25	25	25	25	25	25	

TABLE 1B (GIRLS)
Chronological Age

M.A.	10	11	12	13	14	15	16	17	Total
8	1								1
9	1	5	1						7
10	3	5	7	1					16
11	7	4	2	2	1				16
12	3	7	3	2	5	4			24
13	2	1	4	2	3	7	4		23
14			3	4	3	8	7	4	29
15		3	2	6	3	5	5	6	30
16				5	1		4	5	15
17			1	2	7	1	2	6	19
18			2	1	2		2	3	10
19							1	1	2
									192
	17	25	25	25	25	25	25	25	

As a method of study, scores from the separate tests were correlated with the Stanford-Binet mental age for both sexes. Correlations next were made between the remaining tests. Lastly, partial correlations were made between the various tests with mental age held constant.

RESULTS

Table 2 shows the results of the correlations between the battery of tests. Correlations were made separately for both

sexes, and with one exception the correlations agree quite closely.

TABLE 2

Inter-correlations	Boys	Girls
1. Healy P.C. II and (a) Binet M.A.	.51	.50
(b) Ferguson Boards	.52	.45
(c) Trabue Completion	.42	.41
(d) Scaled Information	.50	.44
2. Ferguson Boards and (a) Binet M.A.	.51	.54
(b) Healy P.C. II	.52	.45
(c) Trabue Completion	.33	.44
(d) Scaled Information	.49	.49
3. Trabue Completion and (a) Binet M.A.	.71	.65
(b) Healy P.C. II	.42	.41
(c) Ferguson Boards	.33	.44
(d) Scaled Information	.65	.64
4. Scaled Information and (a) Binet M.A.	.79	.78
(b) Healy P.C. II	.50	.44
(c) Ferguson Boards	.49	.49
(d) Trabue Completion	.65	.64

The highest positive correlations are found between the Stanford, Trabue, and Information. These tests give a regular distribution, relatively few scores being disproportionately high or low on the latter two. The present findings of the Trabue are in close agreement with previous studies. On the original scale, Trabue (2) obtained a coefficient of .74 with the Stanford-Binet on a group of 50 boys and girls, ages 10–16, and a higher coefficient of .85 on a smaller group of 39 boys in the same age range. Worthington (3), working with a slightly dull group of 123 children, found a correlation of .73 between Trabue Completion and Stanford mental age. Shulman, using the present modified form, reports from an unpublished study, a similar correlation of .75 for a group of 255 boys whose IQ's range from 80 through 110.

No correlations between the Scaled Information test and the

Binet have been published, as far as we know. Since the Scaled Information gives a fairly high correlation with the Trabue, which in turn correlates well with the Binet, one might expect the close relationship between mental age and general information found in this study.

The literature on Healy Pictorial II in relation to Stanford-Binet is more extensive. Numerous investigators of the Healy test have found a low but positive correlation between the two. Thus Vernon (4) obtained a correlation of .54 on a group of 79 subjects 15 years or older. Morgenthau (5), taking a larger number of 116 subjects, obtained exactly the same correlation. Worthington (3) found a lower correlation of .41 with 143 unselected cases. Shulman's correlation of .50 obtained from 225 boys with IQ's between 80–110 corroborates the findings of this study. Bronner (6) who obtained low correlations, by the rank-difference method, between Binet IQ and P.C. II score on an age group ranging between 11 and 17 years concludes that "Healy Pictorial Completion Test II correlates but slightly with intelligence quotients obtained on the Stanford-Binet scale."

The Ferguson Form Boards, like the Pictorial Completion II, show low positive correlations with Binet mental age. Ferguson (7) in his original article found a correlation of .81 between test score and grade on 364 subjects ranging from first grade through college levels. Correlations between Army Alpha results, estimated intelligence, and class standing of 36 sixth grade subjects were .51, .50, and .56, respectively. In a recent article on the Ferguson Form Boards, Wood and Kumin (8), correlating form board score with Binet mental age, give correlations of .55 for 180 boys with IQ of 110 or above, and a somewhat lower correlation of .42 for 121 girls with the same IQ range. They also obtained correlations of .54 and .53 between the same tests with a group of 243 boys and 161 girls with IQ's ranging from 80 through 110.

It is interesting to note that the Ferguson Form Boards and Pictorial Completion II show correlations almost identical to those obtained when each test was correlated separately with

mental age, although the coefficients are too low to be of much significance. In comparing Ferguson and P.C. II with the Completion Sentences and Information Test, we find in most instances only a slight degree of relationship. This is further verified by holding mental age constant and correlating the four tests. The resulting partial coefficients, given in table 3, show a negligible relationship between the various tests when intelligence in terms of Binet mental age is not a factor.

TABLE 3

Partial correlations (Binet M.A. held constant)	Boys	Girls
1. Healy P.C. II and (a) Ferguson Boards	.35	.35
(b) Trabue Completion	.09	.13
(c) Scaled Information	.19	.09
2. Ferguson Boards and (a) Trabue Completion	−.05	.14
(b) Scaled Information	.17	.11
(c) Healy P.C. II	.35	.25
3. Trabue Completion and (a) Scaled Information	.21	.27
(b) Healy P.C. II	.09	.13
(c) Ferguson Boards	−.05	.14
4. Scaled Information and (a) Healy P.C. II	.19	.09
(b) Ferguson Boards	.17	.11
(c) Trabue Completion	.21	.27

The highest partial correlation is again between the form boards and P.C. II. There is thus some indication that a closer relationship exists between these two tests than might ordinarily be suspected, even though the correlations and partial correlations for the Ferguson and Healy tests are somewhat low. Also the type of ability that the Stanford-Binet measures seems to be less of a factor in these two tests than it is in the Completion Sentences and the Information test.

SUMMARY AND CONCLUSIONS

Of the four tests studied in this article, a test for general information was found to have the highest correlation with intelligence as measured by Binet mental age.

One test for language apperceptions, a modification of the Trabue Completion Scale, also showed a significantly high correlation with Binet mental age.

The Trabue Completion and Information tests in turn were found to have a moderately high positive correlation with one another.

Pictorial Completion II and Ferguson Form Boards give low positive correlation with Binet Mental Age, as well as with the Information and Sentence Completion tests. The present findings corroborate the findings of previous investigators in this respect.

We may thus conclude that two of the tests studied, namely, the Trabue-Kelly-Spivak Language Completion test and the Lowe-Shimberg Scaled Information, may be used in a clinical test group to substantiate the rating on the Stanford-Binet, as well as to give additional information regarding the individual. Healy Pictorial Completion II and Ferguson Form Boards, on the other hand, evidently test abilities relatively independent of those measured by the Stanford-Binet, and are primarily valuable for supplementing the clinical study.

BIBLIOGRAPHY

1. Bronner, A. F., Healy, W., and Others. A Manual of Individual Tests and Testing. Little, Brown & Co. 1927.
2. Trabue, M. R. Completion-Test Language Scales. Teachers College, Columbia University Contributions to Education, 1916: 77, pp. 78.
3. Worthington, M. R. A Study of Some Commonly Used Performance Tests. J. Appl. Psychol., Vol. X, 1926, No. 2, pp. 216–227.
4. Vernon, P. E. A Study of the Norms and the Validity of Certain Mental Tests at a Child Guidance Clinic. Brit. J. Psychol., Vol. VII, 1937, pp. 115–137.
5. Morganthau, D. R. Some Well-Known Mental Tests Evaluated and Compared. Arch. of Psychol., Vol. VII, 1922, No. 52, pp. 54.
6. Bronner, A. F. The Apperceptive Abilities of Delinquents. J. Delinq., Vol. VII, 1922, No. 1, pp. 43–54.
7. Ferguson, G. O. A Series of Form Boards. J. Exper. Psychol., Vol. II, 1920, pp. 47–58.
8. Wood, L., Kumin, E. A New Standardization of the Ferguson Form Boards. J. Gen. Psychol., 1939, 54, 265–284.

NORMS OF BUSINESS COLLEGE STUDENTS ON STANDARDIZED TESTS: INTELLIGENCE, CLERICAL ABILITY, ENGLISH

W. LESLIE BARNETTE
Pace Institute, New York City

CHECKING back through the Psychological Abstracts and the Psychological Index, the writer finds almost no reports relative to intelligence norms, as well as other types of scores, on business college students relative to standardized tests. This report represents a beginning in remedying this deficiency.

PREVIOUS REPORTS

Triplett (3) is the only report this writer located in the psychological journals. The 1930 Thurstone (ACE) examination was given to 200 commercial college students and the Otis SA Higher examination to 400 students. His median on the ACE corresponded to the 48 percentile for college students, while 45 per cent of the commercial group equalled or excelled the college median. The commercial group was only 2 and 3 points (men and women respectively) below the Otis college median. No critical ratios are cited. Husband (5, p. 29) reports the "chemical-commerce" group at the University of Wisconsin has a median intelligence percentile of 52.5. The particular test involved here is not stated. Andrew and Paterson (1) administered the Minnesota clerical test to university business students, but cite no norms in the appendix. This is likewise the case in the recent manual of directions for this test (2). Dodge (3, pp. 27–37), using clients from the Adjustment Service of New York, cites norms on adult workers

carefully selected as to their occupational success. His medians differ from those of Andrew and Paterson (2).

TESTS AND RESULTS

Pace Institute offers courses in preparation for the major fields of business. These include accounting, advertising and marketing, selling and marketing, credits, and secretarial practice. Professional courses are also offered to prepare the student for the qualifying entrance examinations in these fields. All courses of study are approved by the Education Department of New York State. The secretarial practice course involves 2 years of study; courses in the fields of marketing, advertising, selling, and credit science all involve 3 years; accountancy, 3 and 4 years, the latter being the CPA preparation course. All matriculated students must have completed high school.

We began giving standardized tests in the fall of 1937. At the inception of that program, admittedly slight, we were primarily concerned about the general standing of our students when compared to the typical college population. At the outset we wanted primarily to know whether our student body was above or below the median college freshman level.

Tests given were the Thurstone (ACE) examination (1937 and 1938 editions), the Minnesota Vocational Test for Clerical Workers, and the Iowa English Training (Series IB). Table I presents test norms and critical ratios (D/PE_D) comparing our scores with the various test norms of the Thurstone and Iowa examinations. The ACE 1937 norms were taken from the full report published in the *Educational Record* (7). The complete score tabulations for the 1938 edition were not available at the time of writing. Consequently the preliminary norms (8) have been used for such comparisons.[1] Norms for the Iowa English test have been taken from the manual of directions (6).

[1] Tentative norms show insignificant deviations from the complete tabulated norms cited in the *Educational Record* for April, 1939. The writer has therefore not re-calculated the critical ratios herein.

Negative signs before the critical ratios indicate our medians are lower than those quoted by the test authors; positive signs indicate our scores are higher. The formulae used have been taken from Garrett (4): the PE_{mdn} is formula No. 32 (p. 206), the PE_D is No. 38 (p. 215).

All data on scores are computed on matriculated freshmen. Such students carry 20 hours of work per week. All students who come to the school who have had college experience of one term or more are excluded from the tabulations. Likewise, students carrying only a few courses have been omitted. A few cases of matriculated students carrying 16 or 17 hours weekly

TABLE 1

Norms and Critical Ratios: ACE 1937 and 1938, Iowa English

N	Test	Group	Mdn	Q_1	Q_3	D/PE_D
			Thurstone, ACE 1937			
158	Completion	Pace	28.0	21.50	36.63	− 3.05
10749		ACE	30.32	22.16	39.24	
158	Arithmetic	Pace	22.60	10.0	33.0	+ .23
10740		ACE	22.33	12.99	33.84	
158	Artif Lang	Pace	24.70	12.40	32.0	− .01
10277		ACE	24.71	16.05	37.25	
158	Analogies	Pace	31.80	21.90	37.90	− 3.0
10607		ACE	34.18	27.50	39.64	
158	Opposites	Pace	52.10	40.10	61.20	− 2.1
10561		ACE	54.35	41.82	68.09	
158	Gross Score	Pace	154.80	122.10	178.90	−35.0
68999		ACE	167.08	128.67	208.87	
			Thurstone, ACE 1938			
132	Q-score	Pace	23.56	16.99	36.34	− 3.0
8652		ACE	26.85	18.70	33.90	
132	L-score	Pace	60.0	48.0	70.27	− 3.90
8652		ACE	64.79	51.38	77.02	
132	Gross Score	Pace	82.70	68.55	98.25	− 4.5
10494		ACE	90.28	71.05	107.36	
			Iowa English (ET, IB)			
158	Gross Score	Pace	103.60	79.40	120.0	+ 3.2
10803		Iowa	97.0	72.0	123.0	

have been included. Of 290, the total number of scores tabulated for both ACE 1937 and 1938, there are only 18 such individuals included.

It will be seen from a perusal of the above table that most of the differences are not statistically significant (D/PE_D less than 4.0). The one large exception to this is the gross score on the ACE 1937. The reader will note the addition here of over 50,000 cases on the published test norms, resulting in a small PE_{mdn}. Had the total number of cases here been of the same magnitude as on the sub-test scores, this difference would not have been so augmented, although it would still be well beyond the statistical requirement of a true difference. The other example of a significant difference is the gross score on the 1938 edition. The conditions under which the 1937 examination was administered were by no means as favorable as those of the succeeding year. The 1937 edition was given after the close of the regular school day. Suitable desks for the students were not available. The 1938 examination was given under a better-controlled classroom situation. The writer believes this may, in part, account for the smaller difference in favor of the college group on the 1938 examination. Nevertheless under these better conditions of administration, the difference in medians is still significant.

The above data show business students are not inferior to college students in English training. In this case the tendency is towards superiority of our group, while on the intelligence examinations the tendency is towards inferiority.

Table 2 presents the data on the Minnesota clerical examination. The Minnesota norms are taken from the revised manual of directions (2). The norms here, as well as in the earlier publication (1), are presented only in terms of deciles. Q_1 and Q_3 were computed by the writer: $Q_1 = \frac{30\text{–}20 \text{ percentile}}{2}$, $Q_3 = \frac{80\text{–}70 \text{ percentile}}{2}$. The Dodge norms are taken from his study on occupational ability patterns (3). Clients of the

Adjustment Service were used here. Although unemployed at the time of testing, Dodge selected his groups to fulfill certain success criteria: at least three years' experience in the particular field, at least one year employment under a single employer, and longer experience in this occupation than in any other. On the whole, his norms for special occupational groups are lower than those quoted in the Minnesota manual. All critical ratios cited in this table represent comparisons between the median reported for the particular occupational group and the median for our students.

TABLE 2

Norms and Critical Ratios: Minnesota Clerical

		Test 1: Number Checking				Test 2: Name Checking			
		Mdn	Q_1	Q_3	D/PE_D	Mdn	Q_1	Q_3	D/PE_D
	Men								
113	Pace	108.2	91.7	119.0		106.7	91.1	122.2	
	Minnesota								
500	Adult empl	83.0	62.0	102.5	+13.3	78.0	54.0	111.5	+11.8
120	Empl clerical	135.0	117.5	156.5	− 9.9	126.0	108.5	147.5	− 6.7
29	Accts, bkps	144.0	129.5	161.0	− 9.0	127.0	113.0	144.0	− 5.0
	Dodge								
46	Accountants	128.9	111.3	148.1	− 5.6	133.3	105.0	154.2	− 5.5
54	Bookkeepers	125.6	105.0	145.6	− 4.5	128.6	107.5	147.2	− 6.0
	Women								
53	Pace	111.6	98.1	123.6		118.9	103.7	133.8	
	Minnesota								
232	Adult empl	109.0	92.0	133.0	+ .9	111.0	86.5	137.0	+ 2.4
284	Empl clerical	144.0	124.0	162.0	−12.5	152.0	130.5	175.5	−10.7
181	Stenog	147.0	129.0	165.0	−12.9	158.0	141.5	174.0	−13.0
	Dodge								
50	Secretaries	117.0	97.5	138.1	− 1.2	130.0	117.0	147.9	− 3.0
50	Stenog	125.0	107.2	145.0	− 3.4	140.9	123.9	155.0	− 6.0
50	Office clerks	121.4	101.0	136.5	− 2.6	123.3	95.0	140.7	− 1.0

With one exception, all of the above test norms with which our student scores are compared represent scores of adult

clerical workers. No student norms seem to be available. It will be seen that men students are significantly superior on both number and name checking to men adults employed. Our women students show a tendency in this same direction, but the differences are not statistically significant. The writer believes this is the fairest comparison to be made as regards student scores. Comparing our scores with Minnesota's norms for employed clerical workers, our students are significantly inferior to all groups on both tests (men and women employed clerical, men accountants and bookkeepers, women stenographers). The writer feels that to use such groups as a comparison-basis in deciding superiority or inferiority of students is unwarranted. All differences are not statistically significant when Dodge's norms are used. Our men students are uniformly inferior on both tests to his groups of accountants and bookkeepers. With our women students, five out of the six differences are not statistically significant with his special occupational groups. This one exception to the general tendency here relates to women stenographers on name checking. On the whole the data show men business students are superior to adult gainfully employed men on both tests, whereas women students tend in this direction. Men students are uniformly inferior on both tests to all clerical workers. The same is true in the case of women students, although the tendency here is not as pronounced.

CONCLUSIONS

1. On all sub-test norms of the 1937 and 1938 ACE examinations, business school students are not significantly inferior to university students. The tendency is, however, towards inferiority on the majority of the sub-test medians. This is more pronounced on the L-score of the 1938 edition (D/PE_D of -3.90), the chances being practically 100 in 100 that this is a significant difference in the direction of inferiority.

2. Business school students are significantly inferior to the typical college group when gross scores of these two groups are

compared on both these examinations. The difference is not as striking on the 1938 edition. It is suggested that the better conditions under which this latter edition was administered is a factor here.

3. On the Iowa English Training business students are equal to college students. There is a slight tendency, statistically unreliable, for the former to be superior.

4. Men students are significantly superior on the Minnesota clerical test (number and name checking) to adult gainfully employed men. There is a tendency in this direction for women business students, although the differences are not statistically reliable.

5. Men business students are significantly inferior to all employed clerical as well as special groups of clerical workers (accountants, bookkeepers) on both parts of the Minnesota test.

6. A similar finding is reported for women business students when they are compared with Minnesota's norms for adult clerical workers. The differences are not so striking when Dodge's norms are used from his study on clients from the Adjustment Service. In only one of six instances is there a statistically significant difference indicating inferiority.

7. There is a need for norms based on business students for both parts of the Minnesota clerical test.

REFERENCES

1. Andrew, D. M. and Paterson, D. G. Measured characteristics of clerical workers. Bulletin of Employment Stabilization Research Institute. Vol. 3, No. 1. Minneapolis: University of Minnesota Press. 1934.
2. Andrew, D. M. and Paterson, D. G. Manual of directions for Minnesota Vocational Test for Clerical Workers. New York: Psychological Corporation. 1938.
3. Dodge, A. F. Occupational ability patterns. Teachers College Contributions to Education #658. Bureau of Publications, Teachers College, 1935, 97 pp.
4. Garrett, H. E. Statistics in Psychology and Education. New York: Longmans, Green and Co. 1938, 2nd edition, 485 pp.
5. Husband, R. W. Applied Psychology. New York: Harper and Brothers. 1934. 644 pp.

6. STODDARD, G. D. Manual of directions for Iowa Placement Examinations, Series I, Revised, Forms A and B. Iowa City: University of Iowa Press.
7. THURSTONE, L. L. AND T. G. The 1937 examination for college freshmen. Educ. Record 1938, 19, 209–34.
8. THURSTONE, L. L. AND T. G. Preliminary report of norms for the 1938 psychological examination for college freshmen. American Council on Education. 1938 (mimeographed report).
9. TRIPLETT, R. J. Intelligence of commercial college students. Pers. J. 1934, 13, 92–93.

NEWS AND NOTES

From June 20 to July 31, 1940, will be held the fifteenth annual session of the Vassar Institute of Euthenics, for the study of human relations and the application of the arts and sciences to the betterment of human living. The announcement, with the above heading, lists the lecturers and discussion leaders for this ambitious and well established program. The list includes college presidents, a professor of philosophy, a curator of ethnology, a professor of English, of botany, of economics, of political science, of physiology, a pediatrician, a psychiatric consultant, two exponents of progressive education, a professor of biblical literature, but only two psychologists. At least, only two of the names are listed as members of the American Psychological Association.

It is a significant comment on psychology or psychologists, we hardly dare say on the rest of the world, to see a project devoted to "the study of human relations and the application of the arts and sciences to the betterment of human living" in which psychology has so small a representation.

THE EDITORS

Owing to the present war conditions the Twelfth International Congress of Psychology, which was planned to take place at Edinburgh, Scotland, in July, 1940, has been postponed though not abandoned. The invitation of the City and University of Edinburgh is not withdrawn, and the Local Organization Committee hope that the Twelfth Congress may still meet in Edinburgh when circumstances permit the continuation and completion of the preparation. The Local Organization Committee is composed of Professor James Drever, President, and Professor Godfrey Thompson, General Secretary, Moray House, Edinburgh.

Dr. S. J. Beck, Head of the Psychology Laboratory of Michael Reese Hospital, will offer a course on The Rorschach Method in Personality Study and Clinical Diagnosis for five days, from June 24–28, 1940, inclusive. The course will include teaching the technique of administering the Rorschach Method, scoring the responses, and interpretation with especial emphasis on clinical classification. Response records as obtained from various healthy personality groups and from clinical groups (schizophrenia and some neuroses to be included), will be scored, analyzed and interpreted. Primary aim of the course will be to demonstrate the test's

practical application in investigating the whole personality, with particular reference to its clinical use. Interested persons may receive further information by addressing the Medical Librarian, Michael Reese Hospital, 2908 Ellis Avenue, Chicago, Ill.

This Journal is pleased to publish the following note from Dr. Gardner Murphy, Department of Psychology, Columbia University, New York City: "Many Chinese psychologists, having lost all their books, journals, reprints, etc., through bombing, are desperately in need of any psychological printed matter which American psychologists care to spare. Through the kindness of friends who are going to China early in the summer, I am undertaking to make a shipment of books and journals to western China to which several of the east China universities have moved and are still carrying on their work under extreme difficulty. Another shipment I hope will be sent in the fall. Expenses of shipment will be taken care of." Those wishing to make donations may communicate directly with Dr. Murphy at the above address.

The Educational Policies Commission of the National Education Association has recently published a book by Dr. John K. Norton, Teachers College, Columbia University, entitled *Education and Economic Well-Being in American Democracy*. During the three and a half years of preparation of this volume, many economists and other experts have assisted the Commission in reaching its conclusions and the policies outlined. To secure the greatest economic effects of education, the Commission proposes: first, the expansion of the American school system to provide every child with a *minimum* of ten years' schooling while increasing the *average* number of years' schooling from ten to fourteen; second, further development of carefully planned vocational education, along with general education, to give more youth the necessary preparation for productive jobs; and third, provision of this education on an effectively free basis so that all youth can go as far in school or college as their talents permit. National progress and educational opportunity tend to go along together, the report points out.

BOOK REVIEWS

ALAN E. TRELOAR. *Elements of Statistical Reasoning.* New York: John Wiley & Sons, 1939. x, 291 pp. $3.25.

This work represents a welcome departure from the customary "methods" treatment of statistics, in that the primary emphasis is placed upon the logical foundations of the statistical approach to a wide range of problems rather than upon the manipulation of statistical techniques. For the most part, the discussion centers around the nature of the problems and the philosophy of the statistical method. In most instances this is supplemented by one or more of the customary formulae that are convenient for calculation.

The general field embraced by this work includes the major statistical problems encountered in psychological research—the analysis of the characteristics of frequency distributions, correlation, proportions and probability, and sampling errors. Scanty attention is devoted to the problems of collecting data and graphic representation, and no mention is made of time series analysis.

The book is intended by the author to be an elementary treatment. This objective appears to have been attained in the sense that almost no mathematical and no previous statistical knowledge is a prerequisite. However, the author's experience with graduate students of serious scientific interest and background appears to have raised the general tone of the book to a distinctly advanced level. The terminology is precise but often foreign to the beginning student. Furthermore, the treatment of some of the more complex concepts—such as correlation surfaces—appears to be too brief for the novice. In view of the depth of the treatment of numerous aspects of correlation, the omission of a discussion of curvilinear relationships and of multiple correlation mars this otherwise excellent portion of the book.

In brief, the work is not a substitute for the orthodox "methods" texts but is a commendable supplement. For the advanced mature student, this work may prove to be a good introduction to the subject, and should materially broaden his interpretation of statistical results.

DONALD E. CHURCH,
Ohio University

GRABBE, PAUL, in cooperation with GARDNER MURPHY. *We Call It Human Nature.* Harper and Brothers, New York, 1939.

From the first glance at the outside to the last written word and drawing inside, Grabbe's book is different from other texts for general psychology. It reads like a collection of fairy stories and is illustrated copiously with appropriate pictures and drawings. To one in a rut it will lack something which has been regarded heretofore as essential for a sound treatise on psychology, namely, selected bits from illustrative experiments. In fact, it is possible that some may compare it with a correspondence course in navigation. But the comparison could more profitably be reversed upon the traditional wordy texts.

Grabbe has discovered a method for presenting psychological facts in an appetizing form of understandable words and pictures. The usual topics are presented in the different chapters. At first glance one is inclined to regret that so much space has been given to pictures and drawings; but upon scrutiny, one recognizes that they are perhaps more valuable than the reading material. The instructor who wants to be the dominant stimulus should not use this book.

The book is admirably suited to a first course in psychology in the high school and as a supplement to the regular text in the first course in college.

J. R. GENTRY,
Ohio University

H. MOORE. *Psychology for Business and Industry.* New York: McGraw-Hill Book Co., 1939. Pp. xi + 527.

As far as the content of Professor Moore's book is concerned, it does not differ very much from the books on applied and industrial psychology already on the market. There are the customary sections on vocational guidance and selection, testing and test construction, job analysis, promotion, training, motivation, accidents, fatigue, advertising, and selling. What makes this book distinct is its approach to the different problems discussed. Professor Moore has succeeded in projecting himself into the position and point of view of the enlightened businessman, and has thus gained an insight into the ways of business and industry which is very apparent in his book.

Another noteworthy characteristic of Professor Moore's book, in line with the point of view referred to, is its inclusion of a vast amount of experimental research on the various problems of business gleaned from the personnel departments of business concerns and the research centers maintained by outstanding industrial concerns. The time has passed for psychologists to close their eyes to the research and investigations carried on outside the immediate precincts of the universities. It is no accident, for instance, that Dr. Gerhard Domagk is the seventh industrial research worker to obtain the Nobel Prize. Professor Moore is keenly aware of the work being done "outside," and the student of business and indus-

trial psychology, reading Professor Moore's book, will benefit from this awareness.

The business of psychology, Professor Moore points out, is human nature as revealed in specific situations. In business and industry the most important of these situations is the relation between the man and the job. Man must work if he is to live. More than that, if he is to lead a happy, well-adjusted life, he must do the type of work for which he is best suited by nature and nurture. This being the case, man must be careful to choose the kind of job that will fit him. Looking at this problem from the employer's point of view, man must possess the qualities which the available job demands. The hiring of men who are not suited to their jobs involves an unwarranted loss in money and time.

The obvious solution of this problem is guidance for the employee, and selection on the part of the employer. Professor Moore very ably reviews the methods that are in present use to effect such guidance and selection. He shows how they are constructed, how they are applied, and what the result of their application is. On the whole, this result is gratifying.

Bringing employer and employee together is, however, only the first step towards a smoothly running business or industrial organization. The initial good will, established between the employer and employee, has to be maintained. This is effected by an intelligent study of such problems as job satisfaction, boredom, fatigue, and output.

Professor Moore enhances the value of this book also by presenting, for illustrative purposes, samples of tests, rating scales, and other devices actually in use in business and manufacturing plants. The tables, graphs, and figures with which the book is richly studded lend authenticity to what Professor Moore has to say. The reader feels that the author of this book is not a mere compiler of information gleaned from books and periodicals, but actually is in close contact with the business and industry of which he is writing.

The topics of report, and the required and supplementary reading suggested by Professor Moore at the end of each chapter are of added significance to both teacher and student. The book will make an invaluable addition to such already well established works as Husband's Applied Psychology, Myer's Industrial Psychology, and Viteles' Industrial Psychology.

PETER HAMPTON,
University of Manitoba

HILGARD, ERNEST R. and MARQUIS, DONALD G. *Conditioning and Learning.* New York: D. Appleton-Century Company. 1940. Pp. xi + 429. $2.75.

To all too many psychologists, busy with the day's work and their own research interests, the need to gain a better grasp of the vast literature on

conditioning is matched by time limitations making it impossible to do so, and when one has read extensively, but perhaps not too intensively, there remains the perplexing problem of just what to do with it all. For example, how is one to find his way between such a view as that set forth in Smith and Guthrie's *General Psychology in Terms of Behavior*, even though he has kept up with events since that book appeared, and the view set forth in Robinson's *Association Theory Today?* Or again, is the conditioned response the unit of habit, or is conditioning rather to be regarded as a source of deductive principles? Certainly the need for a thorough review of the literature is great. But it must be more than a review. To be of greatest value it must be a critical exposition that does not facilely answer all the questions, but gives the reader a better basis for integration of his knowledge through placing the problems in their proper historical setting and through relating them to each other and to closely allied problems in the field of learning. This was the need, and the book *Conditioning and Learning* has met it admirably.

In the preface the writers state, "Theories of learning which make use of conditioning principles are not related in a simple manner to the facts from conditioning experiments." This sentence, perhaps more than any other, provides a key to the writers' attitude toward conditioning. Although their own research studies in this field have been extensive and significant, they are not so carried away with the importance of conditioning as to see in it a simple solution of any major psychological problem. In the first chapter, entitled *The Place of Conditioning in Psychology*, a curious paradox is mentioned. It is said by the writers that the data of experimental psychology by methods other than that of conditioning have always been the richer in content, in the fields of both learning and perception. "Why," they ask, "was a new method hailed as a sign of the birth of an objective science of psychology before that method had been thoroughly tried out?" In attempting to answer this question, the whole movement is treated historically, with especial reference to its place in psychology in this country.

Following two historical chapters, seven chapters are devoted to a discussion of the experimental and theoretical literature on various phases of conditioning in its relation to learning. It is this relationship which constitutes the major emphasis of the book as the title would suggest. The remaining chapters are entitled respectively Problem Solving, Voluntary Action, Personality, and Neurophysiological Mechanism of Conditioning. Even in these chapters it need hardly be said, learning is an important part of the story.

Each chapter is followed by a series of notes, typically 3 pages in length, giving much additional reference material. Of their bibliography, which contains 973 references, the authors state that it is not meant to be com-

plete, but is fairly so for the years following the publication of Razran's (1937) extensive bibliography. The book also contains a glossary, especially devoted to terms specific to conditioning, and including translations of Pavlov's terms.

This is not a book which admits of an easy and conventional evaluation. While the content is primarily factual, the writers' own points of view inevitably appear at times. With them some may wish to quarrel. Such disputes are hardly for the garden variety psychologist, as only those who are working and thinking intensively in this field will be in a position to challenge the authors' interpretations. The book is written with the thoroughness and discernment that has characterized the earlier work of the writers. That it will be indispensable to students of conditioning and learning goes without saying. That it will be read widely by psychologists and laymen who read seriously in the field of psychology seems likely. That it will do more than provide a readily available source of information, modifying and clarifying thinking of many psychologists is to be hoped. Just how important a book it is can hardly be said at present. To this reviewer it seems to be one of the few really significant books in the field of psychology to have appeared in recent years.

AMOS C. ANDERSON,
Ohio University

NEW BOOKS AND PAMPHLETS

New books and pamphlets for review should be sent to James P. Porter, Editor, JOURNAL OF APPLIED PSYCHOLOGY, Ohio University, Athens, Ohio.

The American Management Association has recently published the following bulletins:

Attitudes and Emotional Problems of Office Employees. HAROLD B. BERGEN AND OTHERS, Office Management Series No. 87; 50 cents; 34 pp.

Employee Selection. GUY W. WADSWORTH AND MARIAN A. BILLS.

Salary Administration. L. C. LOVEJOY AND N. D. HUBBELL. Office Management Series No. 88; 75 cents; 40 pp.

Office Equipment and Record Keeping. ROBERT W. WYTHES AND OTHERS. Series No. 89; 75 cents; 39 pp.

Progressive Office Management Policies. F. L. ROWLAND AND OTHERS. Series No. 86; $1.00; 44 pp.

Conditioning and Learning. ERNEST R. HILGARD AND DONALD G. MARQUIS. D. Appleton-Century Company, New York, 1940. $2.75. 429 pp.

How Character Develops. FRITZ KUNKEL AND ROY E. DICKERSON. Charles Scribner's Sons, New York, 1940. $2.50. 274 pp.

How We Learn. BOYD HENRY BODE. D. C. Heath & Company, Boston, 1940. 308 pp.

The Invasion from Mars. HADLEY CANTRIL. Princeton University Press, Princeton, N. J. 1940. $2.50. 228 pp.

Minimum Essentials of the Individual Inventory in Guidance. GILES M. RUCH AND DAVID SEGEL. U. S. Department of the Interior, Vocational Division Bulletin No. 202. Office of Education, Washington, D. C. 15 cents. 83 pp.

Proceedings of The American Catholic Philosophical Association, Vol. XV. CHARLES A. HART, Editor. Association of the Catholic University of America, Washington, D. C. 291 pp.

Psychometric Tables for Obtaining the Vapor Pressure, Relative Humidity, and Temperature of the Dew Point. C. F. MARVIN. U. S. Department of Agriculture, Government Printing Office, Washington, D. C., 1937. 10 cents. 87 pp.

Statistical Analysis in Educational Research. E. F. LINDQUIST. Houghton Mifflin Company, 1940. $3.00. 266 pp.

A Study of a Group of Children of Exceptionally High Intelligence Quotient in Situations Partaking of the Nature of Suggestion. RACHEL MCKNIGHT SIMMONS. Bureau of Publications, Teachers College, Columbia University, New York, 1940. $1.60. 112 pp.

Washburne Social-Adjustment Inventory. JOHN N. WASHBURNE. World Book Company, Yonkers, New York, 1940. Package of 25, $1.30.